

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/88809>

Copyright and reuse:

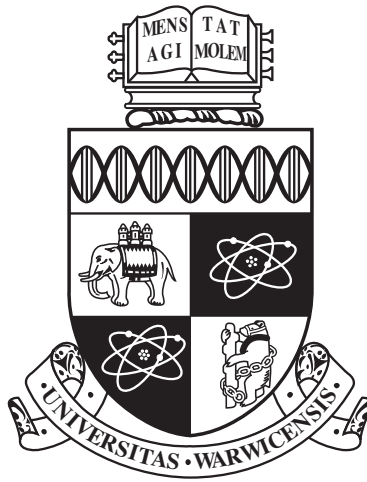
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Transcriptional analysis of salt shock in
Brassica oleracea

by

Christine Patricia Hicks

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Systems Biology

December 2016

THE UNIVERSITY OF
WARWICK

Dedicated to Nan..

Contents

List of Tables	ii
List of Figures	iii
Acknowledgments	iv
Declarations	v
Abstract	vi
Chapter 1 Introduction	1
1.1 Environmental uncertainty and effects on agriculture in the 21st Century .	1
1.2 Salinity in agriculture	2
1.2.1 Effects of high salinity on plants	3
1.2.2 Tolerance mechanisms against high salinity	5
1.2.3 Regulation of salt tolerance mechanisms	9
1.3 Methods of gene expression analysis	15
1.3.1 Microarrays	16
1.3.2 RNAseq	17
1.3.3 Time-series analyses	19
1.3.4 Transcriptomic analyses of salt stress in the literature	19
1.4 Polyploidy in the plant kingdom	20
1.5 C-genome <i>Brassica</i> species	21
1.5.1 Polyploidy and the Brassicaceae	22
1.5.2 C-genome Diversity Fixed Foundation Set	24
1.6 Aims of the project	26
Chapter 2 Materials and Methods	28
2.1 Plant growth	28
2.1.1 Growing <i>Brassica</i>	28
2.1.2 Experimental designs	28

2.1.3	Stress conditions	32
2.1.4	Harvesting and storage of leaf material	33
2.1.5	Phenotype measurements	33
2.2	Laboratory methods	34
2.2.1	RNA extractions	34
2.2.2	qPCR	35
2.2.3	Microarrays	36
2.2.4	RNAseq: library preparation and sequencing	38
2.3	Bioinformatics methods	38
2.3.1	MAANOVA	38
2.3.2	Gaussian Process Two Sample (GP2S)	39
2.3.3	Multiple Dataset Integration (MDI)	39
2.3.4	Causal Structural Inference (CSI)	40
2.3.5	RNAseq: data analysis	40
2.3.6	Creating and querying a BLAST database	44
2.3.7	GO term enrichment analysis	44
2.3.8	Phylogenetic tree construction	44
2.3.9	Determining diurnal expression	44
Chapter 3 Development of transcriptomic resources in <i>Brassica oleracea</i>		45
3.1	Chapter overview	45
3.2	Stress response pathways are activated in <i>B. oleracea</i> GD33DH	46
3.3	Preparing a stress-specific transcriptome of <i>B. oleracea</i>	54
3.3.1	RNAseq analysis of <i>B. oleracea</i> GD33DH under abiotic and biotic stress conditions	55
3.3.2	Generation of a non-redundant transcriptome database	56
3.4	Microarray probe design	56
3.5	Annotating the <i>B. oleracea</i> TO1000 genome with GO terms	58
3.6	Discussion	63
3.7	Chapter Summary	64
Chapter 4 High-resolution time series transcriptomics of salt shock in <i>Brassica oleracea</i>		65
4.1	Chapter Overview	65
4.2	Results	66
4.2.1	Physiological effects of salt shock in <i>B. oleracea</i> GD33DH	66
4.2.2	Experimental design of the time-course microarray experiment	67
4.3	Processing high-resolution time-series transcriptomic data	70

4.3.1	Data transformation and normalisation of <i>B. oleracea</i> GD33DH time series gene expression data using MAANOVA	70
4.3.2	Quality Assessment and LOWESS Transformation	71
4.3.3	Fitting a mixed model to gene expression data	73
4.3.4	Extracting predicted gene expression data from MAANOVA	74
4.4	Analysis of high-resolution time-series transcriptomic data	74
4.4.1	Identifying differentially expressed transcripts	75
4.4.2	Determining time of differential expression	82
4.4.3	Establishing a chronology of the early salt shock response in <i>B. oleracea</i> GD33DH	84
4.4.4	Transcripts involved in ABA biosynthesis are up-regulated following salt shock	88
4.4.5	Modelling ABA signalling following salt shock	88
4.4.6	Hormone cross talk in the regulation of the salt shock response . . .	91
4.4.7	Differentially expressed ion transporters	92
4.4.8	Differential expression of transcripts mapping to key genes with known involvement in abiotic stress responses	94
4.4.9	Differentially expressed transcription factor families	97
4.4.10	The effect of salt shock on circadian regulated transcripts	102
4.4.11	Clustering differentially expressed transcripts	102
4.4.12	Validation of differentially expressed transcripts using RNAseq data	107
4.4.13	Inferring regulatory networks	111
4.5	Discussion	113
4.5.1	Clustering based on time of first differential expression	114
4.5.2	Clustering based on expression profiles	118
4.5.3	Validation of results	119
4.5.4	Limitations of array and experimental design	119
4.5.5	Further work	120
4.6	Chapter Summary	121

Chapter 5 Analysis of gene expression in response to salt shock in wild

	C-genome <i>Brassica</i> species	122
5.1	Chapter overview	122
5.2	Results	123
5.2.1	Preliminary salt shock phenotype screen of wild C-genome <i>Brassica</i> species	123
5.2.2	Differential responses to salt shock in wild C-genome <i>Brassica</i> . . .	131

5.2.3	Second salt resilience screen with Doubled Haploid C-genome wild <i>Brassica</i> lines with selection of lines for sequencing	131
5.2.4	Experimental design	133
5.2.5	Selection of susceptible and tolerant DH lines	136
5.2.6	Mineral analysis of selected tolerant and susceptible DH lines	137
5.2.7	Transcriptomic analysis of selected tolerant and susceptible lines . .	138
5.2.8	Analysing genes differentially expressed in C10128 responding to salt shock	140
5.2.9	Analysing genes differentially expressed in susceptible and tolerant lines	143
5.2.10	Transcription factors regulating salt shock response have genotype specific expression patterns	151
5.2.11	Over-represented TF families in salt shock susceptible and tolerant <i>Brassica</i> lines	154
5.2.12	Comparing gene expression of susceptible and tolerant lines to the DHSL150 rapid cycling parent	155
5.3	Discussion	157
5.3.1	Experimental limitations and further work	161
5.4	Chapter Summary	161

Chapter 6 Transcriptional divergence of stress responsive paralogs in C-genome *Brassica* species 163

6.1	Chapter overview	163
6.2	Results	164
6.2.1	Analysis of sub-genome contribution in <i>Brassica</i> species under salt shock	165
6.2.2	Analysis of sub-genome contribution under various stress conditions in <i>B. oleracea</i> GD33DH	167
6.3	Investigating the importance of genes expressed in multiple copies in the transcriptional response of <i>Brassica</i> under stress conditions	168
6.3.1	Paralogous genes expressed in response to salt shock in <i>Brassica</i> species	170
6.3.2	Paralogous gene expression in other stress conditions in <i>B. oleracea</i> GD33DH	172
6.4	The transcriptional fate of genes expressed in multiple copies	174
6.4.1	The fate of genes expressed in multiple copies in <i>Brassica</i> species in response to salt shock	175

6.4.2	The fate of genes retained in multiple copies in GD33DH in response to stress conditions	176
6.5	Discussion	178
6.5.1	Sub-genome dominance	179
6.5.2	Stress-specific expression of genes in multiple copies	179
6.6	Chapter summary	180
Chapter 7	General discussion	182
7.1	Development of transcriptomic resources in <i>Brassica oleracea</i>	183
7.2	High-resolution time series transcriptomics of salt shock in <i>Brassica oleracea</i>	183
7.3	Analysis of gene expression in response to salt shock in wild C-genome <i>Brassica</i> species	185
7.4	Transcriptional divergence of stress responsive paralogs in C-genome <i>Brassica</i> species	186
Appendix A	MAANOVA analysis script	187
Appendix B	DESeq2 analysis script	190
Appendix C	Quality control of GD33DH RNAseq reads	192
Appendix D	Transcript quality control of <i>de novo</i> assembly	194
Appendix E	Differentially expressed transcription factors in time-series experiment	196
Appendix F	Cluster plots	199
Appendix G	GO terms associated with clusters	203
Appendix H	Quality control of wild <i>Brassica</i> species RNAseq reads	207
Appendix I	Additional data files	209

List of Tables

1.1	The composition of the <i>B. oleracea</i> TO1000 genome	24
2.1	qPCR primer sequences	36
2.2	Microarray hybridization plan	37
3.1	Metrics for genome guided assembly of RNAseq reads	54
3.2	Assembly metrics for <i>de-novo</i> assembly of RNAseq reads	55
3.3	Species included in the pan-Brassicaceae BLAST database	58
4.1	Mineral analysis of <i>B. oleracea</i> GD33DH leaves in early stages of salt shock	67
4.2	The number of differentially expressed transcripts identified by <i>F</i> -tests, followed by different methods of FDR correction	75
4.3	Ion transporter families	93
4.4	Cluster size summary following MDI clustering	104
4.5	Validation of selected genes in time series experiment and RNAseq experiment	109
5.1	Summary of Cg-DFFS lines used in the second diversity study	133
5.2	Doubled Haploid lines selected for RNAseq	136
5.3	Differentially expressed genes in four DH <i>Brassica</i> lines following salt shock treatment	140
5.4	Differentially expressed genes in C10128 following salt shock treatment . .	141
6.1	Sub-genome origin of differentially expressed genes in <i>Brassica</i> lines following salt shock treatment	164
6.2	Sub-genome origin of differentially expressed genes in <i>B. oleracea</i> GD33DH following different stress treatment	167
6.3	The number of multiple-copy genes in the differentially expressed genes of the <i>Brassica</i> lines under salt shock	170
6.4	The number of multiple-copy genes in the differentially expressed in GD33DH under different stress conditions	172
E.1	Top up-regulated TF families	197

E.2	Top down-regulated TF families	198
G.1	Over-represented GO terms of clusters	206

List of Figures

1.1	Response to high salinity in plant roots and shoots.	6
1.2	Phylogenetic tree of Brassicaceae and related rosid outgroups	20
1.3	Triangle of U	21
1.4	Evolution of the <i>Brassica</i> genome	23
1.5	Development of the <i>Brassica</i> C-genome Diversity Fixed Foundation Set . .	25
2.1	Design of time-series experiment described in Chapter ??	29
2.2	Design of wild <i>Brassica</i> S1 experiment described in Chapter ??	30
2.3	Design of wild <i>Brassica</i> DH experiment described in Chapter ?. Paired treatment and control plants were grown side-by-side to control for variation in glasshouse conditions.	31
3.1	Investigation of the effects of different stress treatments on <i>B. oleracea</i> GD33DH	47
3.2	Experimental validation of the RD26 marker gene following stress treatment in <i>B. oleracea</i> GD33DH	52
3.3	Number of probes per <i>B. oleracea</i> TO1000 gene ID present on the C-genome <i>Brassica</i> microarray	57
3.4	BLAST hits of TO1000 gene sequences	59
3.5	GO annotation of the TO1000 genome	61
3.6	Top 20 level 5 GO annotations of the Biological Process category.	62
4.1	Loop design to link biological replicates over time	68
4.2	Loop design to link treatment and control conditions	69
4.3	RI plots before and after the application of global and regional LOWESS transformation	72
4.4	Venn diagram showing the number of transcripts with significant differential expression using <i>F</i> -Tests in MAANOVA	76
4.5	Gaussian Process Two Sample test on time series expression data	79
4.6	Top 20 differentially expressed transcripts	81
4.7	Z-indicator plots and corresponding transcript expression profile	83

4.8	Differentially expressed transcripts at each time point	85
4.9	Selected over-represented GO terms over 24h of transcripts differentially expressed after salt shock in <i>B. oleracea</i> GD33DH	86
4.10	ABA biosynthesis and signalling	89
4.11	Casual Structure Identification network inference of ABA signalling compo- nents	90
4.12	Time of first differential expression on selected groups of transcripts	92
4.13	Expression profiles of differentially expressed <i>B. oleracea</i> GD33DH tran- scripts whose orthologs have previously reported functions in the abiotic stress response in other plant species	95
4.14	Differentially expressed TF families	98
4.15	Diurnally expressed transcripts following salt shock	103
4.16	Plots of the mean gene expression profile of a selection of clusters of co- expressed transcripts differentially expressed after salt shock in <i>B. oleracea</i> GD33DH	105
4.17	Correlation between difference between treatment and control in time series experiment and logFC of gene expression in salt shock RNAseq data	107
4.18	Inferred network models using the Causal Structure Identification algorithm	112
5.1	Phylogenetic tree of wild C-genome <i>Brassica</i> species (S1 lines) used in the study	124
5.2	Effect of salt shock on selected wild C-genome <i>Brassica</i> species	125
5.3	Natural variation of resilience to salt shock in wild C-genome <i>Brassica</i> species (S1 lines)	129
5.4	Clustering wild C-genome <i>Brassica</i> species based on response to salt shock	132
5.5	Effect of salt shock on selected wild C-genome <i>Brassica</i> S1, DH and cultivates lines	134
5.6	Resilience measurements of lines selected for RNAseq	136
5.7	Mineral analysis of a tolerant and a susceptible Doubled Haploid <i>Brassica</i> line	138
5.8	PCA loading plots of the RNAseq count data	139
5.9	Venn diagram to illustrate the overlap of up-/down-regulated differentially expressed in response to salt shock in a susceptible and tolerant line of wild C-genome <i>Brassica</i> species	143
5.10	Expression of a selection of hormone related genes in salt shock susceptible and tolerant <i>Brassica</i> species	145
5.11	Expression of a selection of differentially expressed antioxidants genes in salt shock susceptible and tolerant <i>Brassica</i> species	147

5.12	Expression of a selection of differentially expressed genes involved in photosynthesis and respiration in salt shock susceptible and tolerant <i>Brassica</i> species	148
5.13	Selection of differentially expressed ion transporter genes in salt shock susceptible and tolerant <i>Brassica</i> species	150
5.14	Selection of differentially expressed TFs in salt shock susceptible and tolerant <i>Brassica</i> species	151
5.15	Enrichment analysis of selected TF families in susceptible and tolerant <i>Brassica</i> lines following salt shock treatment	153
5.16	Complete linkage hierarchical cluster analysis of salt shock induced changes of gene expression in three lines of <i>Brassica</i>	156
6.1	Log ₂ expression of sub-genome expression in <i>Brassica</i> species in response to salt shock	166
6.2	Log ₂ expression of sub-genome expression in <i>B. oleracea</i> GD33DH following different stress treatment	169
6.3	Log ₂ expression of genes expressed in multiple copies in <i>Brassica</i> species in response to salt shock	171
6.4	Log ₂ expression of genes expressed in multiple copies in <i>B. oleracea</i> GD33DH in response to stress	173
6.5	Gene fate of 2-3 copy paralogous genes in <i>Brassica</i> species	175
6.6	Expression profiles of triplet paralogs with different fates during salt shock in GD33DH	177
6.7	Gene fate of 2-3 copy paralogous genes in <i>Brassica</i> species	178
C.1	Quality assessment of RNAseq reads	193
D.1	Sequence distribution of <i>B. oleracea</i> GD33DH <i>de-novo</i> assembly and sources of contamination	195
F.1	Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in <i>B. oleracea</i> GD33DH	200
F.1	Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in <i>B. oleracea</i> GD33DH	201
F.1	Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in <i>B. oleracea</i> GD33DH	202
H.1	Quality assessment of RNAseq reads	208

Acknowledgments

I would like to thank my supervisors Prof. Vicky Buchanan-Wollaston and Dr. Jay Moore, for all their help and guidance throughout my PhD project. I would also like to express my thanks to Systems Biology for providing my funding.

I am extremely grateful to Dr. Katherine Denby, Dr. Peter Walley, Dr. Justyna Prusinska, Dr. Krzysztof Polanski, Dr. Christopher Penfold, Dr. Yi-Fang Wang and Lesley Ward, amongst others for their help and guidance with the scientific and technical aspects of my project. Thank you to members of my advisory panel, Jim Beynon, Guy Barker and Sascha Ott for meeting regularly to discuss my project.

I am thankful to all members of the C030 office, past and present, especially Claire Stoker, Sarah Harvey, Stephanie Kancy, Jack Grundy, Adam Talbot, Krzysztof Polanski, Lennie Foster and Philip Law for moral support. Finally, I would like to thank friends and family, for all their encouragement and support throughout my PhD project.

Declarations

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work in this thesis has been undertaken by myself except where otherwise stated.

Abstract

Keeping the global population fed in times of climate change and population growth is considered to be one of the greatest challenges of the 21st Century. Plant stress is defined as any external factor that negatively impacts on growth, productivity, reproductive capacity or survival. The use of salinized water in agriculture is likely to become a more regular occurrence, as diminishing freshwater supplies are available for crop irrigation. High salt drastically affects growth and it is therefore necessary that crops be bred to be able to withstand such adversity.

Recent advancements in technology allow us to measure gene expression on a genome wide scale, techniques resulting in the development of theoretical models of regulation and the identification of key regulatory genes have been used in *Arabidopsis*. There is need to transfer this knowledge from model plant to crop, ensuring the application of such technologies to the issue of food security.

A large microarray experiment was performed during this project in which the expression of over 60,000 genes were measured in *Brassica oleracea* GD33DH over a period of 36 hours following salt shock. The use of bioinformatics tools allowed the identification of 7,141 significantly differentially expressed genes in the early response to salt shock in GD33DH. Additional information on the time of differential expression revealed potential genes and mechanisms indicating that metabolism was highly affected by salt shock.

Germplasm from crop wild relatives in breeding programmes is a crucial source of genetic material to replace variation lost through years of selective breeding allowing the development of crops with higher stress tolerance. By screening a collection of wild C-genome *Brassica* species for salt shock tolerance, tolerant germplasm was identified and sequenced alongside susceptible germplasm. Comparative analyses revealed the genes and mechanisms used by wild *Brassica* species protect themselves from the adverse effects of salt shock.

Whole genome duplication events occurring in the recent evolutionary history of C-genome *Brassica* was examined whereupon it was found that stress specific duplicate genes are on average expressed more highly than single copy suggesting that WGD has implications on the response to stress.

These results provide a wealth of potential gene targets for future study and germplasm that can be used in the development of stress tolerant *B. oleracea* varieties.

Chapter 1

Introduction

1.1 Environmental uncertainty and effects on agriculture in the 21st Century

Food security is a key challenge of the 21st century: ensuring that the world's population has access to sufficient calories and a secure food supply regardless of fluctuations in production and price throughout the year. Recent advances in agriculture have resulted in a fall in the number of under-nourished people since 1990 despite significant population growth. There are still, however, 795 million people in the world without adequate nutrition (FAO, 2015). As people in developing countries become wealthier, they tend to adopt a western-style diet, increasing their consumption of meat, fish and dairy products. This calls for extra inputs and further adds to the pressures on agriculture (Godfray et al., 2010). The global population is expected to reach 9.6 billion by 2050 (United Nations, 2014). This makes the issue of food security even more relevant.

Fluctuating food supplies are compounded not only by population growth but also by climate change (reviewed in Snyder et al., 2009; Wheeler and von Braun, 2013). The relationship between agriculture and climate change is complex. In 2010, agriculture, forestry and land-use change contributed around 20-25% of global greenhouse gas emission (Blanco et al., 2014) in the preproduction phase, through fertiliser manufacture, direct emissions, and post production emissions (Vermeulen et al., 2012). In addition, climate change affects crop production through alterations in atmospheric CO₂ concentration, temperature and water availability (Ahuja et al., 2010). Increased temperatures alter pest populations, causing increased incidences of pests and disease in crops (Rosenzweig et al., 2001). Together, these factors combined ultimately affect crop productivity, leading to reduced yields and higher food prices.

This combination of climate change and population increase means that agriculture needs to adapt to make use of harsher, unpredictable growth conditions (Howden et al.,

2007). Science must play a key role in achieving this aim. Recent advancements in technology mean that data relating to genomes, transcriptomes, proteomes, metabolomes and interactomes can be generated at an unprecedented rate, making biology a ‘big data’ science. These technologies can be used to understand, at a system-level, how crop plants are able to respond and adapt to stress conditions with the ultimate aim of achieving a second ‘green revolution’, thus providing food security for generations to come.

1.2 Salinity in agriculture

Salinization is the accumulation of water-soluble salts in the soil to a level that impacts plant growth and affects crop production. Increased accumulation of salts in the soil can arise from multiple sources including from groundwater supplies, rainfall, rock weathering, sea water intrusions and the use of poor quality water for irrigation (Rengasamy, 2006). Irrigation practices are currently carried out on 20% of total cultivated land, contributing to around 40% of the total calories produced worldwide (AQUASTAT, 2014). The total area of saline soils has been estimated at 397 million ha and it has been estimated that 12 million ha of irrigated land may have gone out of production as a result of soil salinisation (Nelson and Mareida, 2001). Returning this land to productive agricultural use would be highly beneficial in terms of global crop production.

Most important crop species are affected by soil salinity (FAO, 2003). It can cause a 20-50% decrease in maximum yield (Shrivastava and Kumar, 2015) with huge economic implications. It has been predicted that the global annual cost of salt-induced land degradation in irrigated areas could be as much as 27.3 billion dollars (USD) because of lost crop production (Qadir et al., 2014).

The plant response to salt stress is complex and involves the alteration of many biological processes, both physiological and metabolic (Munns and Tester, 2008). Plants vary in their tolerance to salt stress. Adaptive evolution has resulted in two categories of plants based on their salt tolerance. Halophytes e.g. the salt tolerant *Thellungiella halophila*, are capable of managing high levels of soil salinity whilst glycophytes are not sufficiently adapted to grow under high salt conditions (Gupta and Huang, 2014). Most crop plants for instance wheat, *Brassica* species and rice are considered glycophytes and are seriously inhibited by salinity stress conditions (Bernstein et al., 1974).

There are different types of exposure to high salinity that need to be considered. Salt stress is the gradual exposure of plants to salt, whilst salt shock is the application of a high concentration of saline solution to the plant, resulting in a sudden increase in osmotic potential. In the field, farmers are more likely to encounter salt stress rather than salt shock as generally the levels of salt in saline soil do not increase suddenly unless caused by a natural disaster (Shavrukov, 2013).

In the future, salt shock will be more relevant in agriculture, as crops may regularly be grown on salinized soil or reclaimed agricultural lands, such as desert. Irrigation using partially desalinated seawater or brackish water may become a necessary step as freshwater resources diminish (Tester, 2015). An understanding of the molecular mechanisms behind salinity tolerance in crop plants will be crucial for the development of stress tolerant crops that are more capable of maintaining yield and quality despite increased environmental uncertainty.

1.2.1 Effects of high salinity on plants

The effects of high salinity on plants are numerous, ranging from the cellular to the system level and ultimately result in inhibited crop production. Plants respond to high salinity in two distinct phases, the osmotic phase followed by the ionic phase, as defined by Munns and Tester (2008). The time spent in each phase, and the transition period between phases depends on species, timing and severity of the stress (Shavrukov, 2013).

Osmotic phase

The osmotic phase of salt stress induces many physiological changes at the molecular, cellular and whole plant level. These changes include a decreased ability to absorb water from the soil due to high osmotic potential, decreased growth, interruption of the cell membrane, an imbalance of nutrients, excess reactive oxygen species (ROS) production, a decrease in stomatal aperture and a decrease in the rate of photosynthesis (as reviewed in Gupta and Huang, 2014; Munns and Tester, 2008).

When sodium ions (Na^+) in the soil reach above a certain threshold, a hyperosmotic response is initiated in the root within seconds. A reduction of water potential on the outside of the root draws water out, resulting in loss of turgor pressure and wilting of the plant (Downton and Millhouse, 1983; Kumar et al., 2009, reviewed in Khan et al., 2013). Turgor is generally regained within an hour (Shabala and Lew, 2002), indicating that rapid response mechanisms are involved in restoring turgor pressure following salt stress.

Cell expansion and growth are immediately down-regulated in the osmotic phase of salt stress. In the leaves of *Hordeum vulgare* within seconds after adding salt to the roots, the leaf elongation rate decreased to close to zero, and then recovered to levels of 46% and 70% of the non-stress levels within minutes and days, respectively (Fricke et al., 2006). Also, leaf area of *Brassica juncea* seedlings was significantly decreased by up to 70% following 10 days of severe salt stress (Ranjit et al., 2016). In *Arabidopsis thaliana* ('Arabidopsis'), expression levels of cell cycle genes decreased immediately upon addition of salt, and were gradually restored during the recovery period (Burssens et al., 2000; West et al., 2004). This reduction in leaf area and growth, could be an adaptive mechanism in

which soil moisture is preserved through reduced transpiration thus preventing further concentration of Na^+ ions in the soil.

Oxidative stress is caused by excess ROS in the cell. ROS are free radical species (including O_2^- , superoxide anion radical; H_2O_2 , hydrogen peroxide and OH^- , hydroxyl radical) that play a dual role in the response to salt stress. When the metabolic status of a cell is changed, for instance the rate of photosynthesis is reduced and photorespiration is increased then excess ROS are formed. ROS also act as key signalling molecules under stress conditions, regulating signalling of stress related gene expression (Miller et al., 2010). Under normal conditions, ROS are maintained in homeostasis by antioxidants such as superoxide dismutase (SOD) and catalases. Under salt stress conditions, levels of ROS rise, increasing the presence of such detoxifying enzymes (Mittler et al., 2004 and reviewed in Das and Roychoudhury, 2014). In maize under high salinity, a selection of genes encoding antioxidant enzymes were increased in both the leaves and the roots, with a greater number of antioxidant encoding genes differentially expressed only in the root. This suggests that Na^+ retention and detoxification occurs mainly in the root, protecting the photosynthetically active leaves against the effects of high salinity (AbdElgawad et al., 2016).

In glycophytes such as *Arabidopsis*, osmotic stress causes stomatal closure, primarily through an increase in abscisic acid (ABA), but also both ROS and Ca^{2+} secondary signalling have been shown to play a role (Allen et al., 2000; Gilroy et al., 2014; Hernandez et al., 2010; Ward and Schroeder, 1994). Closure of the stomata limits CO_2 availability for fixation by photosystem II (PSII). This results in an increase in cyclic electron flow involving only PSI and an increase in non-photochemical quenching of chlorophyll fluorescence which is highly inefficient (Stepien and Johnson, 2008). This response is used by plants under high-light conditions where excess energy is dissipated as heat and is involved in the long term down-regulation of PSII (Muller et al., 2001).

Following salt stress, susceptible genotypes of *B. juncea* (Liu et al., 2011) and perennial grass species (Mittal et al., 2012) showed a greater reduction in photosynthetic capacity compared to tolerant genotypes, indicating the occurrence of greater damage to the photosynthetic machinery. Tolerant cowpea cultivars showed increased activity of proteins involved in photosynthesis and energy metabolism whilst susceptible cultivars did not, suggesting that the rapid re-establishment of photosynthesis is important in tolerance to salt stress (de Abreu et al., 2014). The faster the recovery of photosynthetic capability following salt stress, the greater is the level of plant survival in the long term (as reviewed in Chaves et al., 2008).

Ionic phase

Na^+ toxicity occurs mainly in the shoot where excess Na^+ ions are stored in the older leaves. Should the build up of Na^+ ions in the shoot pass the threshold of tolerance, toxicity is seen with detrimental effects on cellular function. This ionic phase of salt stress takes place after several days or even weeks depending on the tolerance level of the plant and severity of the stress (reviewed in Gupta and Huang, 2014; Shavrukov, 2013).

The high levels of Na^+ in the shoot may be stored in older leaves or sequestered in the vacuole, as discussed below. Should excess Na^+ rise in the cytoplasm, the effects on metabolism are vast. Excess Na^+ ions can inhibit enzyme activity due competition between Na^+ and potassium ions (K^+) for major binding sites in enzymes involved in a diverse array of metabolic process such as enzymatic catalytic reactions, protein biosynthesis and ribosome function (Marschner, 1995).

Salt-induced senescence of older leaves is a major consequence of the long term effects of excess Na^+ and affects plant productivity under saline conditions. Senescence is the process of tissue degeneration and nutrient recycling in order to support the active growth and development in younger parts of the plant (Hortensteiner and Feller, 2002). Salt stress and changes in the hormonal balance of ABA, ethylene and cytokinin within the leaf can promote the onset of senescence in older leaves undergoing ionic stress (Ghanem et al., 2008). The process is regulated by key transcription factors (TFs) such as the Arabidopsis NAC (NAM, ATAF, and CUC) transcription factor ANAC092 and a senescence-associated protein, SAG29. When the *ANAC092* gene was knocked out in Arabidopsis, delayed chlorophyll loss under high salinity was observed (Balazadeh et al., 2010). When *SAG29*, a gene which is highly expressed during senescence, was knocked out, cells exhibited enhanced viability under salt stress conditions (Seo et al., 2010). These studies indicate the importance of delayed senescence in plant productivity, making the process a key target for crop improvement under high salinity and other water related stress conditions (Rivero et al., 2007).

1.2.2 Tolerance mechanisms against high salinity

Plants have developed several mechanisms against salt stress which include tolerance to both the early osmotic phase and the later ionic phase. An overview of signalling and ion homeostasis of plants in the early stages of salt stress is shown in Figure 1.1.

Early signalling of salt stress

Upon contact of the root with salt, a complex Ca^{2+} signature is propagated from root to shoot via membrane bound RESPIRATORY BURST OXIDASE-D (RBOHD) and is assisted by a ROS triggered element (Evans et al., 2016). This signal varies depending

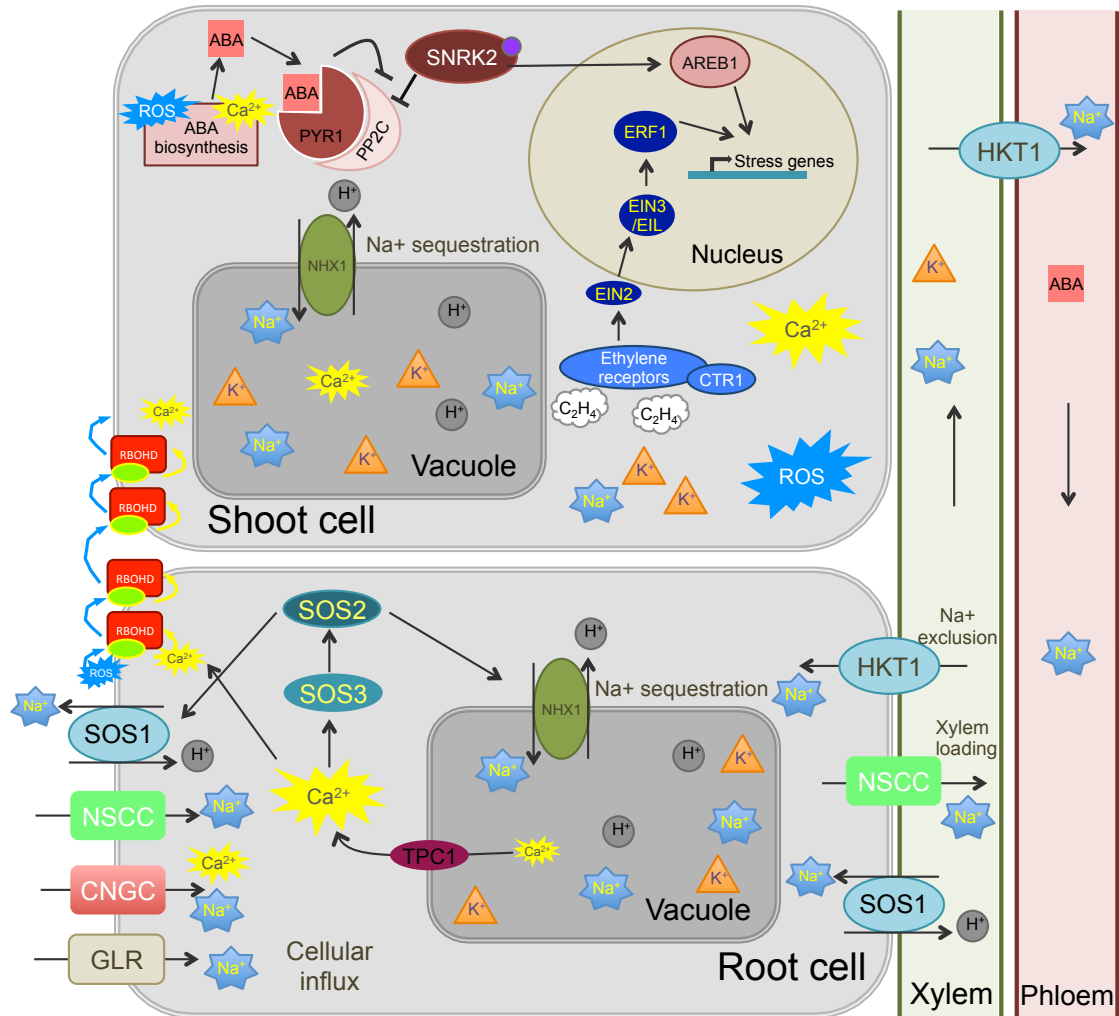


Figure 1.1: Response to high salinity in plant roots and shoots.

An overview of cellular Na^+ transport and signalling components of the salt stress response network in plant roots and shoots. Figure compiled from Evans et al., 2016; Gallie, 2015; Gonzalez-Guzman et al., 2012; Kurusu et al., 2015; Maathuis et al., 2014; Nakashima and Yamaguchi-Shinozaki, 2013; Santiago et al., 2009; Tao et al., 2015.

on species, cell type and severity of the stress (Kiegle et al., 2000; Knight et al., 1997; Tracy et al., 2008). The signal propagation is extremely rapid, 0.37-0.42 mm s⁻¹ in *Arabidopsis* taking around 2 minutes to reach the shoot (Choi et al., 2014). Under non-stress conditions, excess Ca²⁺ is stored in the vacuole. This is released from the vacuole as the signal is propagated through the plant, via Two-Pore Channel 1 (TPC1), which is a slow vacuolar (SV) calcium-permeable ion channel, located on the vacuole membrane (Fig. 1.1). Disruption of *TPC1* through showed a 25 fold decrease in the rate of signal propagation, and the corresponding over-expresser a 1.7 fold increase in signal speed, suggesting an important role in Ca²⁺ signalling (Choi et al., 2014).

This initial Ca²⁺ wave has roles in the up-regulation of ABA biosynthesis in the shoot and in the initiation of secondary signalling messengers such as ROS, which leads to the activation of kinases involved in stress signalling. This ultimately results in down-stream protein activation and transcriptomic changes. Several Ca²⁺ activated TFs have been proposed, including calmodulin-binding transcription activators (CAMTAs) (Pandey et al., 2013) and MYBs (Yoo et al., 2005).

A ROS burst can be generated in a controlled manner by enzymes such as NADPH oxidases and RbohD proteins, to act as a secondary signalling mechanism in response to oxidative stress conditions. ROS and Ca²⁺ signalling work in concert to enable communication between cells and long distance signal propagation in plants via Rboh membrane bound proteins (Evans et al., 2016; Gilroy et al., 2014; Steinhorst and Kudla, 2013).

Salt uptake and transport

As highlighted in Figure 1.1, Na⁺ ions from the soil can enter the root hair cells through non-selective carrier type transporters (such as glutamate receptors (GLRs), cyclic nucleotide-gated ion channels (CNGCs), and other non-selective cation channels (NSCCs) that have yet to be identified), or through ion channels voltage dependent cation channels such as the high affinity potassium transporters (HKTs).

Glutamate receptors are amino acid-activated channels which are permeable to Na⁺, K⁺ and Ca²⁺ (Vincill et al., 2012) and have roles in the response to abiotic and biotic stresses (Forde and Roberts, 2014) in addition to signalling and development (Demidchik and Maathuis, 2007). CNGCs are ligand-gated, Ca²⁺-permeable divalent cation-selective channels and are localised in plasma membrane (Saand et al., 2015). Salt-responsive members of this family are differentially regulated under high salt conditions (Saand et al., 2015) in both the root and shoot (Kugler et al., 2009). It is likely that there are many more NSCCs which have yet to be identified (Maathuis et al., 2014).

The SALT OVERLY SENSITIVE (SOS) pathway, comprising of SOS2, SOS3 and SOS1 and is one of the most well characterised pathways involved in cellular signalling and

homeostasis under salt stress (Zhu et al., 1998) and is highly responsive to Ca^{2+} signalling (Fig. 1.1). The pathway consists of a Ca^{2+} sensor, SOS3, which perceives an increase in Ca^{2+} within the cell and recruits SOS2, a Ser/Thr protein kinase localised to the plasma membrane (Batelli et al., 2007; Liu et al., 2000; Zhu et al., 2007). The complex goes on to activate the down-stream target protein SOS1, which is a Na^+/K^+ antiporter localised in the plasma membrane of the root tip cells and xylem parenchyma (Shi et al., 2000; Wu et al., 1996). SOS1 has two major roles in the tolerance to high salinity. Firstly it functions at the root tip cells to prevent uptake of Na^+ from the soil through extrusion of Na^+ ions. Secondly it acts at the xylem parenchyma, preventing xylem loading to stop Na^+ from reaching the shoot and causing damage to the photosynthetic tissues (Olias et al., 2009).

Once Na^+ ions are in the roots, plants are able to control the long distance transport of Na^+ by preventing loading of Na^+ into the xylem. Many genes are involved in the removal of Na^+ from the xylem, including HKTs, SOS1, members of the CHX cation antiporter family and other non selective cation channels. Should xylem loading occur, Na^+ is transported from root to shoot where it can have detrimental effects on cellular processes (Maathuis et al., 2014).

The high affinity potassium transporter (HKT) gene family plays a key role in the exclusion of Na^+ from the shoot by providing Na^+ selective transport (class I HKTs) and Na^+/K^+ co-transport (class II HKTs) (Maser et al., 2001). These ion transporters are found in xylem parenchyma cells and are involved in preventing xylem loading and subsequent transport of Na^+ ions to the leaves (Sunarpi et al., 2005), as shown in Figure 1.1. When the *HTK1;1* (class I) was disrupted in Arabidopsis, a rise in Na^+ accumulation in the leaves, with concurrent reduction in the roots (Xue et al., 2011). Over-expression of *HvHTK2;1* in barley resulted in enhanced Na^+ tolerance (Mian et al., 2011).

The cation/ H^+ exchanger (CHX) transporter family members localise to intracellular and plasma membranes and may have roles in ion homeostasis following salt stress (Chanroj et al., 2011). Members of the gene family have been implicated in osmotic adjustment and K^+ homeostasis in Arabidopsis (Cellier et al., 2004) and also have a role in K^+ homeostasis during development (Evans et al., 2012; Sze et al., 2004).

Levels of Na^+ in the cytoplasm can be kept low by compartmentalising the excess Na^+ ions (and K^+) in the cell vacuole using membrane bound ion transporters such as Na^+/H^+ exchanger 1 (NHX1) which has affinity to both Na^+ and K^+ , with a preference towards K^+ (Bassil et al., 2012, 2011; Jiang et al., 2010). A *nhx1 nhx2* double mutant showed decreased accumulation of K^+ in the vacuole, but greater Na^+ sequestration suggesting that a key ion transporter involved in the flux of Na^+ into the vacuole has yet to be identified (Barragan et al., 2012; Maathuis et al., 2014). Transgenic analysis has shown V-ATPases and H^+ -PPase (e.g. AVP1) to be key in the maintenance of the transmembrane

electrochemical potential at the tonoplast (Hu et al., 2011; Pasapula et al., 2010; Zhou et al., 2010). This allows for the optimal function of vacuolar Na^+/H^+ antiporters such as NHX1 by supplying H^+ for the sequestration of Na^+ ions in the vacuole (Benito et al., 2014).

It is assumed that Na^+ toxicity is the primary effect of salt stress in plants, however, excess Cl^- also has an adverse effect on the cell. As such, several Cl^- transporters have been identified. The chloride channel (CLC) genes, of which there are 7 in Arabidopsis (Ward et al., 2009), play important roles in pH adjustment and salt tolerance. The *CLC-g* gene, a member of the CLC gene family was found to localise to the vacuolar membrane in Arabidopsis (Nguyen et al., 2016a) and over-expression of the maize chloride channel gene *ZmCLC-d* in Arabidopsis resulted in plants that had reduced Cl^- accumulation and were more tolerant to abiotic stress (Wang et al., 2014).

An alternative method of transport of Na^+ ions from root to shoot without passing over a plasma membrane, is via the entry of Na^+ into the apoplastic space through the lateral roots (Faiyue et al., 2010) and is known as ‘bypass flow’. In cases of salt shock, plasmolysis of the root cells may occur, in which the apoplast detaches from the cell wall resulting in leakage of Na^+ into the apoplastic space and rapid transport to the shoot through the bypass flow mechanism.

Bypass flow of Na^+ transport is well studied in rice and can be reduced by the presence of hydrophobic barriers found in the roots of tolerant plants (Krishnamurthy et al., 2014, 2011). Silicon can be applied exogenously to *O. sativa* in order to create artificial hydrophobic barriers in the outer part of the roots and in the endodermis to reduce transport of Na^+ through bypass flow. This correlated with reduced Na^+ concentration in the shoot (Gong et al., 2006).

Role of potassium in salt stress tolerance

Potassium ions are a key micronutrient required by plants for multiple biochemical and metabolic processes. Potassium is an important co-factor in the function of many key enzymes (Marschner, 1995), K^+ gradients are used as a mobile energy resource, to overcome local energy limitation (Gajdanowicz et al., 2011) and are able to alter stomatal aperture (Dietrich et al., 2001; Fischer, 1968).

K^+ and Na^+ are similar physico-chemically as both are monovalent inorganic cations (Benito et al., 2014), however there are fundamental differences between the two. In the response to salt stress, high K^+ levels play a key role in tolerance, whilst excess Na^+ has a toxic effect on the cell. Increased levels of Na^+ in the cytoplasm (over 100mM) interfere with the activity of many enzymes, and those that require K^+ as a co-factor are especially affected (Marschner, 1995; Munns and Tester, 2008).

One of the major salt tolerance mechanisms seen in plants is via the maintenance of a

high K^+Na^+ ratio (Maathuis and Amtmann, 1999). Plants achieve a high K^+Na^+ ratio by two possible methods, either by preventing Na^+ from reaching the shoot (Na^+ exclusion), by increasing K^+ uptake and storage of Na^+ and K^+ in the vacuole (ion sequestration). Many ion transporters, including HKTs and NHXs, are able to transport both Na^+ and K^+ , and even show a greater affinity towards K^+ transport, unless the concentration of Na^+ is higher than that of K^+ (Barragan et al., 2012; Jiang et al., 2010, 2014b; Maathuis et al., 2014; Sunarpi et al., 2005). Thus, it is possible that the sequestration and uptake of excess Na^+ ions is a side effect of the plant attempting to increase and maintain a high concentration of K^+ ions to ensure cellular processes are unaffected by high levels of Na^+ .

Natural variation in the salt tolerance of various *Arabidopsis* ecotypes was associated with the K^+Na^+ ratio (Sun et al., 2015). Potassium fertilization together with the application of salicylic acid to sugar beet crops growing on saline soils showed increases in tolerance and yield (Merwad, 2016). Similarly, exogenous application of K^+ reduced the symptoms of salt stress in *Brassica campestris*, suggesting K^+ application could be a method for improving crop tolerance to high salinity (Umar et al., 2011).

1.2.3 Regulation of salt tolerance mechanisms

Plant hormones are small, organic signalling molecules that regulate a range of cellular responses during growth and development and play a major role in the plant response to stress. Upon perception of environmental or developmental cues hormones act as signalling molecules to activate signal transduction, leading to the control of global gene expression and the elicitation of an appropriate response to the signal received. Salt stress tolerance is mediated through a collection of hormone signals including abscisic acid, auxin, cytokinin and ethylene. It has become evident that these pathways are not initiated in isolation, rather there is a significant amount of cross talk between pathways either antagonistically or synergistically, allowing fine tuning of the response to a stimulus (Atkinson and Urwin, 2012; Cabot et al., 2013; Seki et al., 2002).

Abscisic acid (ABA)

Abscisic acid is the primary hormone involved in the regulation of gene expression in abiotic stresses such as temperature, high salinity, osmotic stress (Suzuki et al., 2016; Yoshida et al., 2014b). Under non-stress conditions, ABA has an important role in growth and development, including embryo maturation, seed dormancy and germination (Finkelstein et al., 1985; Suzuki et al., 2000). Under salt stress conditions, ABA acts as a regulator of stomatal aperture, preventing unnecessary loss of water through the transpiration stream as well as regulating expression of down-stream stress responsive genes (Daszkowska-Golec and Szarejko, 2013; Geng et al., 2013; Lim et al., 2015; Liu et al., 2005; Wilkinson and

Davies, 2002). During cellular dehydration, levels of ABA have been shown to rise 100-fold above basal levels by 12 h, with levels stabilising around 96 h as plants acclimate to the stress (Verslues and Bray, 2004, 2006).

Salinity stress rapidly activates ABA biosynthetic gene expression through a Ca^{2+} dependent phosphorylation pathway. ABA biosynthesis occurs by several enzymatic reactions in which zeaxanthin epoxidase epoxidates zeaxanthin, the products of which are then converted by 9-cis-epoxycarotenoid dioxygenase to form xanthoxin followed by the conversion of xanthoxin to abscisic aldehyde. The abscisic aldehyde is then oxidised into abscisic acid by ABA-aldehyde oxidase (Barrero et al., 2006; Gonzalez-Guzman et al., 2002; Nambara and Marion-Poll, 2005). This pathway has been constructed *in vitro* in plant protoplasts, in which the presence of ABA in combination with the PYR1 receptor, ABA-insensitive1 (ABI1, a PP2C), the serine/threonine protein kinase SnRK2.6 and the ARBE Binding Factor2 (ABF2) TF and resulted in the activation of downstream ABA responsive gene expression (Fujii et al., 2009; Ng et al., -2013; Raghavendra et al., 2010).

ABA is perceived by the PYR-PYL/RCAR (PYR) receptors, of which there are 14 in Arabidopsis (Gonzalez-Guzman et al., 2012). In the absence of ABA, Protein Phosphatase 2Cs (PP2Cs) repress the subclass III of the SNF1-related protein kinase 2 (SnRK2) family. When ABA binds to the PYR receptor, it activates it and the activity of PP2C is inhibited and the release of the SnRK2s from the PP2C complex occurs (Ma et al., 2009; Park et al., 2009; Santiago et al., 2009). The SnRK2s are then able to phosphorylate ABF (ABA-responsive element binding factors) which go onto initiate ABA regulated gene expression by binding to the conserved cis-regulatory ABA Response Elements (ABRE) ((C/T)ACGTGGC) in the promoters of ABA inducible genes (Gonzalez-Guzman et al., 2012; Yoshida et al., 2010).

The master regulators of ABA signalling under drought and osmotic stress are considered to be the AREB1, AREB2, ABF1 and ABF3 bZIP-type AREB/ABF TFs (Yoshida et al., 2014a, 2010). The proteins encoded for by these genes regulate many key salt stress response TFs, including *MYC2* and *MYB2*, whose proteins in turn induce expression of important stress response genes such as *RD22* and *ADH1* (Abe et al., 2003), NACs and other important bZIP TFs (Hickman et al., 2013; Nakashima et al., 2012; Yang et al., 2009). The dehydration response element (DRE) has been found to act in concert with the ABRE motif, positively regulating ABA-mediated responses to abiotic stress conditions (Narusaka et al., 2003).

Auxin

Auxins are important phytohormones which play essential roles in plant growth and development. Indole-3-acetic acid (IAA) is a key member of the auxin family and is responsible for the majority of auxin action in plants. Biosynthesis of auxins occurs by a

tryptophan dependent or independent pathway (reviewed in Mano and Nemoto, 2012).

The Aux/IAA genes are a large gene family of auxin response repressors. Under non-stress conditions, Aux/IAA repressors dimerize with auxin response factors (ARF) activators which are bound to auxin response elements. This dimerization represses the action of the ARFs and prevents them from activating auxin responsive genes (Tiwari et al., 2001). Under high stress conditions, auxin levels increase and auxin binds to the TIR1 receptor in the SCF^{TIR1} complex, releasing the repressor from the ARF activator, where it is then targeted for proteosomal degradation (Dharmasiri et al., 2005; Kepinski and Leyser, 2005). Meanwhile, the removal of the Aux/IAA repressor from the ARF activator results in the subsequent activation of auxin responsive genes which have roles in root architecture and growth (Guilfoyle, 2007; Petricka et al., 2012).

Under salinity stress, levels of auxin are depleted in the root (Dunlap and Binzel, 1996). In Arabidopsis under salt stress conditions, the redistribution of auxin leads to the suppression of lateral root development altering root architecture, ultimately affecting plant growth and development (Petersson et al., 2009; Wang et al., 2009a). Generation of transgenic potato (*Solanum tuberosum* cv. Jowan) overexpressing *AtYUC6*, an auxin biosynthesis gene showed enhanced drought resistance with reduced levels of ROS in the leaves, suggesting a role for auxin in drought tolerance (Kim et al., 2013).

Cytokinin

Cytokinins are urea based chemicals which act as hormones in plants, and are involved in the regulation of plant growth, development and adaptation to environmental stress. There are two active forms of cytokinins - isopentenyladenine and its hydroxylated derivative zeatin, these can form a variety of conjugates allowing the plant to fine tune cytokinin levels (Frebort et al., 2011).

Under salt stress conditions, cytokinin signalling is inactivated by AHKs, AHPs and ARRs (Type A and B; reviewed in Ha et al., 2012). A reduction of cytokinin levels is correlated with reduced growth. Transgenic over-expressers of isopentenyltransferase (*IPT*), a cytokinin biosynthesis enzyme, showed delayed onset of leaf senescence and increased drought tolerance (Gan and Amasino, 1995; Rivero et al., 2007). Manipulation of the levels of cytokinins have a direct effect on salt tolerance, as such they have been the target of studies in a range of species, for example *Brassica napus* yield was improved through transgenic expression of a cytokinin biosynthesis enzyme *IPT* fused with an AtMYB32 promoter (Kant et al., 2015).

Ethylene

Ethylene is a volatile gaseous hormone (C_2H_4) that is most well known for its role in the promotion of fruit ripening through the conversion of starch and acids to sugars (Abeles et al., 1992). Ethylene biosynthesis starts with the conversion of the amino acid methionine to S-adenosyl-L-methionine (SAM) by the enzyme Met Adenosyltransferase. SAM is then converted into 1-aminocyclopropane-1-carboxylic acid (ACC) by ACC synthase (ACS). ACC synthesis increases with high levels of auxins, especially IAA and cytokinins. The final step involves the action of the enzyme ACC-oxidase (ACO) and oxygen (Wang et al., 2002).

Ethylene has key roles in the early salt stress response and the establishment of the acclimation processes through regulation of various stress response pathways. High levels of ethylene in the long term can negatively affect growth and development, leading to reduced yields and eventual death. Thus it is important that ethylene levels are tightly modulated throughout the stress response, ensuring an adequate balance between survival and the ability to recover growth (Tao et al., 2015).

The pathway of ethylene mediated signal transduction includes five functionally and structurally diverse ethylene receptors ETHYLENE RESPONSE1 (ETR1) and ETHYLENE RESPONSE SENSOR1 (ERS1), subfamily I; ETHYLENE-INSENSITIVE4 (EIN4), ETR2 and ERS2, subfamily II), CONSTITUTIVE TRIPLE RESPONSE1 (CTR1), which has roles in stabilising the ethylene receptors and EIN2, a positive regulator of ethylene signalling which activates the EIN3 transcription factor resulting in the biosynthesis of key ethylene response genes (An et al., 2010; Ju et al., 2012). Under non-stress conditions with no ethylene production, the ETR1 receptor interacts with CTR1 and phosphorylates the endoplasmic reticulum membrane bound EIN2, preventing ethylene induced signal transduction. Upon perception of ethylene, ETR1 changes conformation, inactivating CTR1 leading to the dephosphorylation and cleavage of EIN2. The truncated C-terminus of EIN2 is then able to translocate into the nucleus where it prevents the degradation of the EIN3/EILs TF complex (Ju et al., 2012; Qiao et al., 2012; Wen et al., 2012). EIN3/EIL binds to the ETHYLENE RESPONSE FACTOR1 (ERF1) promoter region to initiate down-stream ethylene response genes. In the nucleus, EIN2 is regulated by EIN2-TARGETING PROTEIN1/2 (ETP1/ETP2) mediated protein turnover (Qiao et al., 2009) and EIN3 is regulated by EBF1/EBF2-dependent ubiquitination and degradation (An et al., 2010; Binder et al., 2007; Gagne et al., 2004; Gallie, 2015; Guo and Ecker, 2003; Potuschak et al., 2003).

Ethylene mediated signalling cascades are associated with three major components of the salt stress response. These include the regulation of ROS and ROS scavengers such as SOD and POD, regulation of ion transporters such as HKTs and the regulation of osmolyte biosynthesis enzymes such as the P5CS genes which are involved in proline

biosynthesis (Tao et al., 2015). Up-regulation of genes associated with these pathways enhance salt tolerance through ROS homeostasis and the maintenance of a high K^+/Na^+ ratio (Tao et al., 2015).

Jasmonic acid

Jasmonates, including jasmonic acid (JA) are lipid based hormones with strong roles in biotic stress resistance and abiotic stress signalling (reviewed in Wasternack and Hause, 2013).

When inactive MYC2, a positive regulator of JA signalling, is repressed by the JAZ, NINJA and TOPLESS (TPL) protein complex (Chini et al., 2009, 2007; Pauwels et al., 2010). When JA accumulates in response to stress, the JAZ proteins are recruited by an E3 ubiquitin ligase CORONATINE INSENSITIVE 1 (COI1) to form a complex which is subsequently targeted for ubiquitination. Following degradation of this complex, the NINJA and TPL proteins are released from MYC2 allowing activation of JA responsive genes (reviewed in Wasternack and Hause, 2013).

Salt tolerant tomato varieties were found to have higher levels of JA compared to susceptible varieties under salt stress conditions (Pedranzani et al., 2003) and salt tolerant barley showed up-regulated JA biosynthesis and JA-responsive genes compared to the susceptible line (Walia et al., 2006a). Exogenous application of jasmonates alleviated salt stress symptoms in barley (Walia et al., 2007), soybean (Yoon et al., 2009) and grape (Ismail et al., 2012).

Salicylic acid

Salicylic acid is a phenolic compound primarily involved in the establishment of basal immunity in response to biotic stress, particularly to biotrophic pathogens with roles in abiotic stress. SA biosynthesis occurs through three separate pathways, the shikimic acid pathway, the isochorismate pathway and the most prevalent SA biosynthesis pathway is the phenylalanine pathway. SA is involved in the regulation of many important biological processes such as photosynthesis, antioxidant defence, proline metabolism and plant water relations under stress (reviewed in Miura and Yasuomi, 2014).

Abiotic stress causes an imbalance in ROS production and scavenging, leading to oxidative stress. Glutathione (GSH) and ascorbate (AsA) are an important antioxidants capable to maintaining homeostatic balance of ROS under salt stress. Biosynthesis of these compounds is mediated by SA, and exogenous application of SA to salt-treated *Triticum aestivum* was found to significantly improve salinity tolerance by increasing GSH and AsA levels in the plant (Li et al., 2013). In Arabidopsis, exogenous SA application has also been shown to restore membrane potential and prevent loss of K^+ through a guard

cell outward-rectifying K⁺ (GORK) channel (Jayakannan et al., 2013). Maintaining high K⁺ is a crucial mechanism for enhancing tolerance to salt stress, as discussed previously. Exogenous application of SA to a variety of plant species has been shown to improve salt tolerance in crops including maize (Gunes et al., 2007), spring wheat (Arfan, 2009) mungbean (Khan et al., 2010) and sunflower (Noreen et al., 2009) through enhanced antioxidant activity (as reviewed in Khan et al., 2015).

Hormone crosstalk

Evidence of complex cross talk between hormone signalling pathways is extensive in abiotic stress signalling (as reviewed in Chan, 2012; Seki et al., 2002; Shinozaki and Yamaguchi-Shinozaki, 2000). Crosstalk between signalling pathways is a crucial mechanism for the tight regulation of gene expression that is needed to tailor an appropriate response depending on the type of stress, severity, time of day and duration. Crosstalk between signalling pathways may be synergistic and/or antagonistic and includes the involvement of hormones, transcription factors, Ca²⁺ and ROS. Brief examples of hormone signalling cross talk in response to salt stress include:

- Auxin and cytokinin - proposed to work antagonistically, possibly mediated by gibberellins in root development (reviewed in Petricka et al., 2012).
- Cytokinins and ABA - Three cytokinin receptor histidine kinases (AHK2, AHK3 and AHK4/CRE1) are negative regulators of ABA and osmotic stress signalling, suggesting that cytokinins and ABA work antagonistically to regulate plant adaption to environmental stress (Nishiyama et al., 2012; Tran et al., 2010, 2007).
- JA and ABA - Both synergistic and antagonistic interactions have been seen between these hormone pathways. It has been suggested that JA activated MYC2 may regulate ABA stress responsive genes (reviewed in Kazan and Manners, 2012; Riemann et al., 2015).
- Ethylene and ABA - the TSS2 and TOS1 proteins in tomato have roles in regulating this cross-talk under osmotic stress conditions (Rosado et al., 2006). In addition, ethylene has been observed as a positive regulator of some aspects of ABA action in the regulation of seed dormancy and germination (Arc et al., 2013; Ghassemian et al., 2000)
- JA and ethylene - proposed antagonistic roles in mediating plant defence against biotic stresses. In *Nicotiana attenuata*, the hormones have roles in reduction of local cell expansion and growth after herbivore attack, allowing for the allocation of more resources towards defence mechanisms (Onkokesung et al., 2010).

Due to the complicated nature of hormone cross talk, further research is needed to disentangle these complex, overlapping signalling networks in response to salt stress to build a complete picture of the salt stress response. This would enable the manipulation of these pathways with the aim of improving salinity tolerance (Ryu and Cho, 2015).

Transcription Factors (TFs)

Integration of signalling of the hormone pathways and secondary signalling messengers occurs downstream of signal transduction, at the level of the transcriptome. Complex gene networks are tightly regulated by TF proteins, which are capable of activating or repressing expression of key stress response genes in the correct spatial and temporal manner (Jaillais and Chory, 2010).

TFs bind to target sequences usually in non-coding DNA, upstream of the transcriptional start site such as enhancer or promoter regions and induce conformational changes in the DNA to allow for access of the transcriptional machinery, or by recruiting key proteins to the transcriptional start site (Schwechheimer and Bevan, 1998). Genes with similar functions are likely to have the same target sequence in the promoter to allow for coordinated expression of a collection of genes in response to a particular stimulus. The major TF families involved in the salt stress response include AP2-EREBP, bZIP, bHLH, MYB, NAC and WRKY (Borkotoky et al., 2013; Kilian et al., 2007; Nakashima et al., 2014; Yoshida et al., 2014b).

Many individual TFs have been shown to play a key role in the response to salt stress in plants. The DREB (dehydration responsive element binding) subfamily, part of the AP2-EREBP TF family has long been associated with abiotic stress responses and functions via an ABA-independent pathway, particularly in response to low-temperature and water deficit (Nakashima et al., 2000; Oh et al., 2005). Recent efforts in over-expressing *GmDREB1A* (from *Glycine max* cv. Jinong 27) in transgenic wheat and *SsDREB* from the halophyte *Suaeda salsa* in transgenic tobacco have resulted in enhanced abiotic stress tolerance without adverse effects on yield (Jiang et al., 2014a; Zhang et al., 2015).

ANAC092 regulates salt induced senescence in Arabidopsis; disruption of *ANAC092* resulted in delayed chlorophyll degradation and increased germination rates under high salinity in Arabidopsis (Balazadeh et al., 2010). Two bZIPs - bZIP53 and bZIP1 together have a key role in the co-ordination of C- and N-metabolic reprogramming in Arabidopsis roots under high salinity, with single and double knockout mutants showing partially redundancy with reduced salt tolerance (Hartmann et al., 2015). The WRKY46 TF has been shown to have roles in the regulation of stomatal movement and lateral root development under osmotic/high salinity stress conditions via cross talk between the ABA and auxin hormone signalling pathways in Arabidopsis (Ding et al., 2015, 2014). Finally, HB7 and HB12, which have been shown to have evolved divergently are involved in the

regulation of growth under water stress conditions and are induced under osmotic stress conditions in *Arabidopsis* (Ré et al., 2014).

1.3 Methods of gene expression analysis

Traditionally, the analysis of gene expression was one gene at a time, with much time and effort spent optimising the process. Methods of gene expression analysis have undergone major advances in recent years, meaning that gene expression can be analysed at the genome-wide level rather than on a per gene basis. This is known as ‘transcriptomics’. Analysing gene expression at the transcriptome level has many benefits, including genome-wide knowledge of gene expression levels which can be compared between different conditions to identify a cause or a response, for instance in healthy versus diseased plants or unstressed versus stressed plants. Measurement of gene expression at the transcriptome level can be made by several methods which are described below.

1.3.1 Microarrays

Microarrays are made from a collection of DNA oligonucleotide probes that are complementary in sequence to genes of interest, arranged in a spotted grid on a glass platform. Depending on the number of probes in the assay, multiple arrays can be included on a single glass slide.

Microarrays work on the principle of hybridization of mRNA to complementary oligonucleotide probes. Quantification is either by fluorescence of Cy3-/Cy5-labelled cRNA prepared from mRNA to be analysed (e.g. Agilent microarrays) or by the hybridization of biotinylated cRNA, followed by staining of the array with a fluorescent molecule (streptavidin-phycoerythrin) that binds to the biotin (e.g. Affymetrix).

Using oligonucleotide arrays such as Agilent, samples can either be labelled in a single colour and compared to a reference which is hybridised separately, or conditions can be compared directly by labelling each sample in two different colours and hybridising them to the same array. Following hybridization, a laser is used to excite the dyes, causing them to fluoresce. The level of fluorescence is proportional to the amount of dye that has been incorporated during hybridization. The pattern of hybridisation will appear as a series of coloured dots, providing a quantitative measure of gene expression. Comparisons between samples can determine which genes have been up-regulated, down-regulated or those whose expression remains the same under the different treatments or conditions.

Affymetrix gene chips are hybridized to a single target sample (similar to one colour arrays described above) and have advantages over Agilent arrays in that more probes can be included on a single array, though the oligonucleotide probes are shorter for Affymetrix

arrays (25mers) compared to Agilent oligonucleotide probes (60mers).

Microarrays are considered a high-throughput method of gene expression analysis since genome wide changes can be assessed rapidly and results can be obtained within days. As the technology is older, they are vastly cheaper than other methods of whole transcriptome analysis meaning more samples can be analysed without incurring high costs. In addition, microarrays are able to capture expression of lowly expressed transcripts, such as lowly expressed TFs, which other quantification technologies such as RNAseq may not capture unless sequencing at depth (Labaj et al., 2011). Microarrays, however, do present some limitations (Shendure, 2008):

- To design the complementary oligonucleotide probes, prior knowledge of the sequence of interest is required. When working with non-model organisms this information may not be readily available, in which case it may be necessary to use the genomic information of a closely related species which increases cross hybridisation potential (Davey et al., 2009; Nieto-Díaz et al., 2007).
- When using a two-colour design, there will be additional noise introduced by using two different dyes. The dyes do not behave in the same way during the labelling and imaging process and so robust statistical methods are required for the analysis (Wu et al., 2002). A dye swap is often used to reduce the effects of dye bias (Churchill, 2004; Kerr and Churchill, 2001), although this does not completely overcome the major source of variability that dye bias can cause between replicates (Martin-Magniette et al., 2005).
- The signal produced by microarrays is an analogue signal, which can be more difficult to interpret and is less quantitative than a digital signal.

1.3.2 RNAseq

The development of high throughput next-generation sequencing technologies in the last decade means that many research groups are choosing to sequence the transcriptome rather than using microarrays to quantify gene expression levels (Martin et al., 2013). RNAseq (RNA sequencing), is a powerful method of whole transcriptome profiling in which the whole transcriptome is sequenced simultaneously providing high resolution gene expression analysis at an unprecedented level of detail. RNAseq allows for quantification of transcript abundance (Sirbu et al., 2012), discovery of new transcripts, analysis of alternative splicing and variant identification in a single experiment (Conesa et al., 2016; Love et al., 2014; Trapnell et al., 2012b).

In order to perform RNAseq, the total RNA is extracted from the sample of interest and mRNA is captured using oligo-dT coated magnetic beads. The mRNA is fragmented,

random-primed cDNA synthesis is performed and the resulting cDNA is prepared into a library. The library preparation step consists of end repair, adapter ligation and PCR amplification, after which the resulting libraries are sequenced, usually using a sequence-by-synthesis method to produce both forward (R1) and reverse (R2) reads of usually 100 bp in length which can be reassembled to give a complete snapshot of the entire transcriptome at the point of sampling (Illumina, 2011, 2014). Once the RNA has been sequenced, the reads are either aligned to a reference genome (Trapnell et al., 2012b), or are assembled *de novo*, without the use of a reference genome (Haas et al., 2013). Transcript quantification is carried out by fitting a model to the read counts that align to a gene or transcript to a negative binomial distribution, taking into account over-dispersion of the data (Anders and Huber, 2010; Love et al., 2014).

The Illumina HiSeq is the current choice for many researchers, as only a small amount of mRNA is needed (100ng to 1 μ g), the library preparation is relatively simple and the resulting data is of high quality. The highly popular Illumina HiSeq 2000, released in 2010 is now depreciated, being replaced with the HiSeq 2500. Other Illumina HiSeq platforms include HiSeq 3000, 4000 and the HiSeq X models now available, depending on the analysis required.

Third generation sequencing technology is now available, for example Single Molecule, Real Time (SMRT) sequencing from PacBio. Since there is no amplification step, third generation sequencers are capable of single molecule sequencing in real time. Polymerase is fixed to the base of a well which is surrounded by Zero-mode waveguide (ZMW). The DNA attaches to the polymerase at the bottom of the well and each of the four nucleotides, fluorescence labelled binds to its complementary base and are incorporated, emitting a light pulse which is recorded by a detector at the bottom of the well. This technology is capable of producing reads of around 10,000 bp in length. This would be highly advantageous for assembly of complex genomes such as *B. oleracea* by bridging large gaps and repetitive regions and also for the unambiguous alignment of paralogous transcripts to the reference genome (Bleidorn, 2016). Disadvantages of the technology is a high error rate of 11–15%, though in the future this is likely to improve (Korlach, 2013).

RNAseq has a multitude of advantages over microarrays:

- The signal produced is digital, resulting in reduced level of background (Shendure, 2008; Wang et al., 2009b).
- Prior sequence knowledge, although desirable, is not essential thus discovery of new transcripts, alternative splicing patterns and novel variants is possible giving new transcriptomic insights.
- Higher dynamic range of up to 9,000 fold has been detected, allowing for the detection of more differentially expressed genes at higher sequencing depths (Conesa et al.,

2016; Tarazona et al., 2011)

- Good agreement with qPCR and microarray results, and high level of reproducibility between technical and biological replicates (Nookaew et al., 2012; Sirbu et al., 2012)
- Bioinformatic tools have become very sophisticated in recent years, allowing increased reliability in mapping to reference genomes, particularly for plants whose genomes are often highly duplicated, the identification of large numbers of SNPs which is useful for association studies and analysing alternative splicing.

However, there are some drawbacks in using RNAseq for gene expression analysis. Highly expressed transcripts such as Rubisco are sequenced many times, 75% of the measurement power of RNAseq measures around 7% of the known transcriptome (Labaj et al., 2011) meaning that lowly expressed transcripts may be missed. Sequencing to a greater read depth allows for a larger proportion of the transcriptome to be analysed with diminishing returns as costs increase (Wang et al., 2009b). In addition, as many bioinformatic analyses are computationally intense and huge amounts of data are output, some users may be limited by computing power and storage availability. Transferring large data files produced by RNAseq analysis presents its own challenges.

It is likely, in the future, that the different methods of gene expression analysis complement each other, but the use of microarrays will be gradually phased out as sequencing costs decrease further.

1.3.3 Time-series analyses

Gene expression in response to stress conditions is dynamic. Plants respond to salt stress conditions in different phases, depending on developmental stage, age, length of treatment and severity of stress (Munns and Tester, 2008). They make drastic alterations to the transcriptome under stress conditions, but also small, yet constant adjustments in gene expression occur, dependent on feedback on the cellular process involved. In addition, many genes are under circadian control and their expression and regulation is dependent on the time of day. Temporal fluctuations of hormone levels under non-stress conditions mean plants alter gene expression to prime themselves for different stresses at different times of the day (Grundy et al., 2015; Ingle et al., 2015).

High-resolution time course transcriptome analyses carried out in the model plant *Arabidopsis* have revealed a wealth of information about developmental processes (Breeze et al., 2011), biotic (Lewis et al., 2015; Windram et al., 2012) and abiotic stress responses (Bechtold et al., 2016), as part of the Plant Responses to Environmental STress in *Arabidopsis* (PRESTA) project, funded by the BBSRC and based at the University of Warwick. These experiments were carried out with many time points allowing the

capture and identification of differentially expressed genes which may be missed using a more conventional single time point analysis and involved the development of multiple bioinformatic tools in Arabidopsis that have been used throughout this thesis.

One of the first steps of processing the vast amount of gene expression data is to carry out cluster analysis, in which genes with similar expression profiles are grouped into clusters. This allows for the identification of genes that are co-expressed and also in the dissection of gene regulatory networks as it possible to identify clusters that influence the expression of other clusters (Mason et al., 2016; Polanski et al., 2014). Biological function of genes can be inferred from the clusters allowing the establishment of a chronology of processes occurring at different stages during the time-course. High-resolution time series data is also amenable to the application of modelling algorithms to the data in order to establish key TF regulators of the response (Hickman et al., 2013; Penfold and Buchanan-Wollaston, 2014; Penfold et al., 2012; Penfold and Wild, 2011). Network inference has been used to determine key regulatory genes controlling the response to environmental stressors in Arabidopsis (Barah et al., 2016; Bechtold et al., 2016; Lewis et al., 2015; Windram et al., 2012) and other crop species (Penfold et al., 2012).

1.3.4 Transcriptomic analyses of salt stress in the literature

RNAseq opened up whole transcriptome analyses of non-model crop species, as there is no need of prior knowledge of genomic sequences or a reference genome to design probes, as was necessary for microarrays. Generally, these experiments are of a single or small number of time points and involve differential expression analyses. Several experiment have been carried out in wild cotton species, in which RNAseq with *de-novo* assembly of the resulting transcriptomes revealed differentially expressed genes over 3, 12, 72 and 144 hpt in *Gossypium aridum* under salt salt stress (Xu et al., 2013a). Alternatively, differentially expressed genes between two contrasting *Gossypium* genotypes at two different time points was used to identify genes and their associated pathways which conferred tolerance to salt stress (Peng et al., 2014). RNAseq has also been used to examine expression partitioning of homeologs and tandem duplications in *Triticum aestivum* under salt stress where transcriptome measurements of two different species of wheat with contrasting tolerance occurred at 6, 12, 24 and 48 hpt after salt stress. Finally, drought and salinity experiments in chickpea revealed both genotype and developmental specific responses following RNAseq analysis in genotypes of contrasting tolerance showed differences in the regulation of metabolic pathways and differential transcription factor expression (Garg et al., 2016).

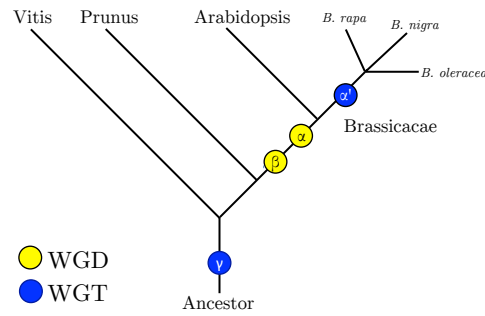


Figure 1.2: Phylogenetic tree of Brassicaceae and related rosoid outgroups

Phylogeny of Brassicaceae and related rosoid outgroups displaying whole genome duplication (WGD, yellow) and triplication (WGT, blue) events. Genome duplication events are named according to the Arabidopsis convention of α' (WGT following *Brassica* speciation event), α (most recent event in the Arabidopsis lineage), β (second most recent) and γ (eudicot paleohexaploidy event). Adapted from Tang and Lyons, 2012.

1.4 Polyploidy in the plant kingdom

Polyploidy, followed by gene loss and diploidization is a major driving force in the evolution of plants. Being polyploid has certain advantages over being diploid, particularly in crop plants. Polyploids often exhibit heterosis, in which traits such as biomass, fertility, yield etc within a polyploid population can exceed that of their diploid counterparts due to a higher level of genetic diversity and heterozygosity (Comai, 2005; Veitia and Vaiman, 2011). The genomes of polyploids are highly plastic, allowing for the divergence of alleles and the evolution of new traits (Roulin et al., 2012).

In the short term, recent polyploids (neopolyploids) suffer from genomic instability in which the cells are unable to cope with the increased amount of genomic material. Chromatids are unable to pair correctly on the meiotic spindle; insertions, deletions and reciprocal translocations lead to the rearrangement and loss of stretches of chromosome (Leitch and Leitch, 2008). After the initial instability, multiple gene copies are retained or lost by several mechanisms, of which some or all may be acting on the gene copies over time.

Through both cytogenetic analysis (Leitch and Leitch, 2008) and genome sequencing (Amborella Genome Project, 2013; Arabidopsis Genome Initiative, 2000) it has been shown that widespread whole genome duplication events (WGD), in which the genome of an organism is doubled or even tripled, have occurred extensively across the plant kingdom. A WGD resulted in the separation of angiosperms (of most dicots, the asterids and rosids) from vascular plants (byrophytes) (Bowers et al., 2003). It is thought that virtually all angiosperms have experienced at least two WGD events in their evolutionary history (Conant et al., 2014; Soltis et al., 2014). WGDs lead to speciation events, particularly in the case of the Brassicaceae (Kagale et al., 2014) in which polyploidy plays an important

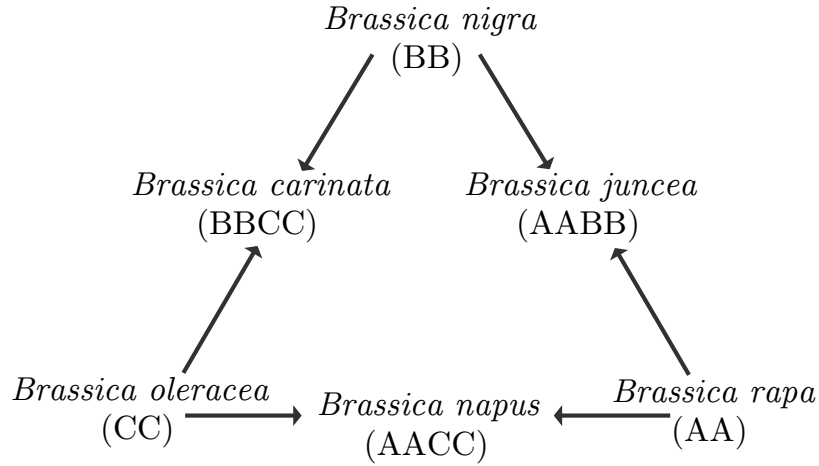


Figure 1.3: Triangle of U

A representation of the Triangle of U, establishing the relationship between key members of the *Brassica* species

role in the vast diversity present within the family (Fig. 1.2).

1.5 C-genome *Brassica* species

The *Brassica* genus belongs to the abundant plant family Brassicaceae and contains many economically important crops such as oil seeds, mustards and vegetables. There are six core members of the genus and their genetic relationship was determined through extensive cytogenetical analyses (U, 1935) and is described by the Triangle of U (Fig. 1.3). Three of the species are diploid (AA genome *Brassica rapa* ($n=10$), BB genome *Brassica nigra* ($n=8$) and the CC genome *Brassica oleracea* ($n=9$)). These diploid species hybridized to produce three allotetraploids, making up the other three species of the *Brassica* genus (AABB genome *Brassica juncea* ($n=18$), AACC genome *Brassica napus* ($n=19$) and BBCC genome *Brassica carinata* ($n=17$)).

Brassica oleracea is an important member of the *Brassica* genus, a vegetable crucifer with a C genome (Fig. 1.3). The species is valued for its diverse morphological features and contains many agriculturally important morphotypes including broccoli, cauliflower, cabbage, Brussels sprouts, kale and kohlrabi. Vegetable *Brassica* have in recent years gained the status of ‘superfood’ because of their association with good health and the presence of high levels of glucosinolates have been found to reduce the risk and progression of cancers (Traka et al., 2013).

The genome sequence of *B. oleracea* was released in 2014 (Liu et al., 2014; Parkin et al., 2014) making genomic analyses of this species more accessible. The TO1000 assembly is 488.6Mb representing around 75% of the predicted *B. oleracea* genome, 92% of the

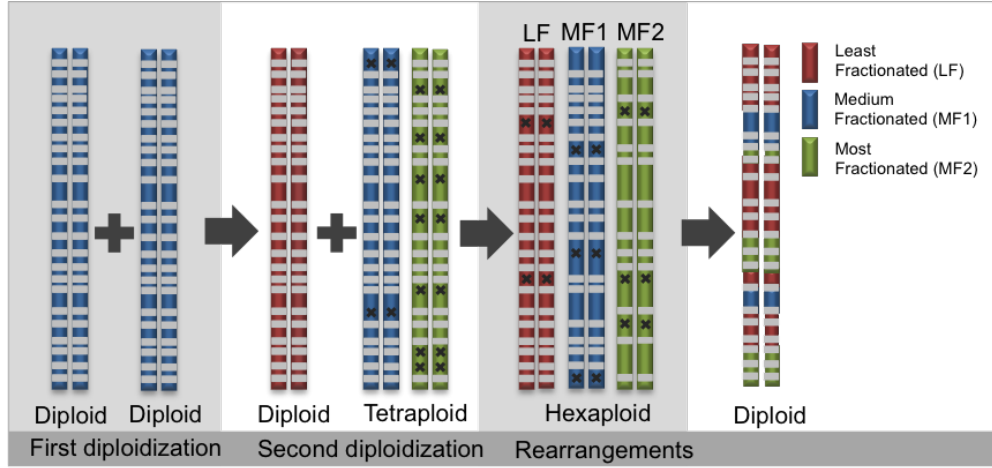


Figure 1.4: Evolution of the *Brassica* genome

A Two Step theory of evolution of the meso-hexaploid *Brassica* genome. A representation of the Triangle of U, establishing the relationships between key members of the *Brassica* species. Rectangles represent genes and black crosses represent genes which are not detectable. The red is the LF sub-genome, blue is the MF1 sub-genome and green is the MF2 sub-genome. Adapted from (Cheng et al., 2012a)

assembled scaffolds are anchored to nine pseudochromosomes. Around 40% of the genome consists of highly repetitive sequence, likely to be uncharacterised transposon related sequences. Genes were annotated using homology to proteins in public databases and also by identifying protein domains *de novo* to give 59,225 gene models (Parkin et al., 2014).

There are 59,225 annotated genes in the current release of the *B. oleracea* genome, 41,174 in the latest *B. rapa* genome and 38,174 gene models in *Raphanus raphanistrum* which is a Brassicaceae that also underwent the WGD events (Moghe et al., 2014; Parkin et al., 2014; Wang et al., 2011b). Given that *Arabidopsis* has ~30,000 genes, it would be expected that following the two WGD events, *B. oleracea* would contained ~90,000 genes, suggesting extensive gene loss following WGD. The number of gene models in *B. oleracea* was 1.4 fold greater than for the *B. rapa* assembly possibly due to the expansion of gene families, tandem duplications, the presence of uncharacterized, repetitive transposon sequences or *B. oleracea* specific genes found in the genome (Cheng et al., 2012b; Parkin et al., 2014; Town et al., 2006).

1.5.1 Polyploidy and the Brassicaceae

Studies have shown that the model *Arabidopsis* has three WGD events in its past history known as paleopolyploidy events; the gamma γ event shared by all dicots and rosids, the α and β events, which have been estimated to be between 170 - 235 MYA (MYA, million years ago) (Bowers et al., 2003) and 50 - 65 MYA (Barker et al., 2009), respectively that are shared by the Brassicales order (Fig. 1.2). This has led to a highly duplicated genome

Sub-genome	# Bo genes	Retained	Paralogous Groups
LF	13,205	Singleton	9,756
MF1	9,281	Duplet	6,984 (13,968)
MF2	7,974	Triplet	2,242 (6,726)

(a) Sub-genome (b) Retained paralogous groups (Bo genes)

Table 1.1: The composition of the *B. oleracea* TO1000 genome

(a) The number of genes originating from the LF, MF1 and MF2 sub-genome in *B. oleracea* TO1000 as determined by Parkin (2014), using data provided in Supplemental Data Table S7. (b) Paralogous groups retained in multiple copies following the Ara-Bra divergence in TO1000 as determined by Parkin (2014), using data provided in Supplemental Data Table S7. The number in parentheses indicates the number of Bo genes within the group (Parkin et al., 2014).

in Arabidopsis which was the first plant genome to be sequenced in 2000 (Arabidopsis Genome Initiative, 2000) and has been extensively further characterised ever since. The Arabidopsis-*Brassica* divergence occurred between 32 and 36 MYA (Town et al., 2006). Following the divergence, a whole genome triplication event occurred (termed the α') occurred in *Brassica* between 24 and 29 MYA (Moghe et al., 2014; Tang and Lyons, 2012) (Fig. 1.2). Following each WGD event, differential gene loss (fractionation) and diploidization followed by extensive genomic rearrangements occurred resulting in the formation of three different genomes within the *Brassica* genus, the A-genome (*Brassica rapa*), the B-genome (*Brassica nigra*) and the C-genome (*Brassica oleracea*) that are able to hybridize to form allotetraploids (Cheng et al., 2012a; Liu et al., 2014; Parkin et al., 2014; U, 1935; Wang et al., 2011b) (Fig. 1.3).

Within the genome, sub-genomes corresponding to each WGD event can be identified by the accumulation of mutations (both synonymous, K_s and autonymous, K_a) and deleterious substitutions over time compared to the ancestral Arabidopsis gene. These sub-genomes have been termed ‘LF’ for the least fractionated with less gene loss, ‘MF1’ for the medium fractionated with moderate gene loss and ‘MF2’ for the most fractionated sub-genome with the most extensive gene loss (Figure 1.4 and Table 1.1a). Genome dominance is the transcriptional dominance of one sub-genome over the other, generally the LF sub-genome shows dominance over the MF1 and MF2 sub-genomes, as fewer mutations have had time to accumulate. Accumulation of transposable elements in older sub-genomes leads to methylation of the DNA and heritable epigenetic silencing of genes (Cheng et al., 2012a, 2016; Parkin et al., 2014; Wang et al., 2011b; Woodhouse et al., 2014).

Gene loss following polyploidy is not random. Particular groups of genes, such as those whose proteins are members of large multi-subunit complexes or highly interconnected gene networks are more likely to be retained so as to maintain the stoichiometric balance

within the genome (Birchler and Veitia, 2007; Conant et al., 2014). Genes that are highly conserved across all eukaryotes, involved in stand alone processes such as essential housekeeping functions, are more likely to remain as single copies (De Smet et al., 2013; Paterson et al., 2006). Transcription factors and genes involved in signal transduction, stress and the circadian clock are more likely to be retained in multiple copies (Blanc and Wolfe, 2004; Jiang et al., 2013b; Lou et al., 2012; Maere et al., 2005a; Parkin et al., 2014; Wang et al., 2011b).

For genes retained in multiple copies, there is opportunity for one copy to diverge whilst another retains the original ancestral function. Genes that are retained in multiple copies have opportunity to diverge in function, whilst retaining an original functioning copy. Possible fates following divergence of multi-copy genes include:

- Pseudogenization - a copy of the gene might either become nonfunctional.
- Neofunctionalization - a copy may acquire a novel function.
- Subfunctionalization - the two duplicates might divide the original function of the gene.

Genes in *B. oleracea* which are retained in multiple copies (either two or three copies) have been teased apart by identifying the ancestral Arabidopsis ortholog for each gene based on protein sequence and assigning it to a sub-genome by K_s analysis (Parkin et al., 2014). It can be seen that many genes (20,694 genes) are retained in multiple copies, whilst comparatively few (9,756 genes) are retained in singular (Table 1.1b).

1.5.2 C-genome Diversity Fixed Foundation Set

The C-genome *Brassica* include *B. oleracea*, a domesticated vegetable crop known for being morphologically diverse which includes broccoli, cauliflower, cabbage, kale, Brussels sprouts and kohlrabi. The evolutionary history of *B. oleracea* is interesting, it has many wild relatives which can be found at various locations across Europe, often at coastal regions (Maggioni, 2015). Reduced level of diversity in the current breeding gene pool, as selection is based on morphological traits, makes breeding for new traits challenging. With the developments in genomic resources, the use of crop wild relatives (CWR) in breeding programs is a popular method of increasing the genetic diversity of commercial varieties without the use of transgenic technologies (Walley and Moore, 2015).

A Diversity Fixed Foundation Set (DFFS) is defined as ‘an informative set of genetically fixed lines representing a structured sampling of diversity across a genepool’. Using a DFFS to introduce natural variation into crop breeding programmes is a method in which the gene pool of a crop species can be widened (Pink et al., 2008; Walley et al., 2012).

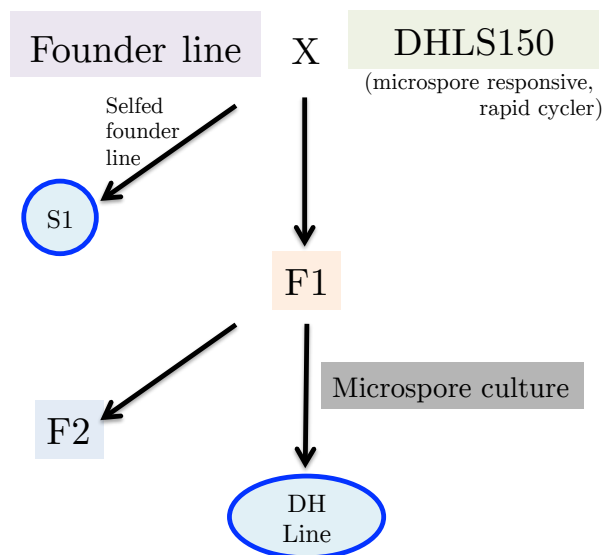


Figure 1.5: Development of the *Brassica* C-genome Diversity Fixed Foundation Set

The work flow of different stages in the generation of the CgDFFS collection. Plant lines circled in blue have been used for experiments in this thesis.

Many DFFS collection have been developed and are made available at the Genetic Resource Unit (GRU) based at the Wellesbourne Campus of the University of Warwick (UK). This includes DFFS collections of carrot, onion, lettuce (Atkinson et al., 2012) and a large range of *Brassica* species. This includes a C-genome DFFS collection (Cg-DFFS), which contains a variety of wild C-genome *Brassica* species, gathered from locations around Europe, in particular from Mediterranean countries, and include *B. incana*, *B. carinta*, *B. hilarionis*, *B. oleracea*, *B. bourgaei*, *B. cretica*, and *B. macrocarpa* (Maggioni, 2015).

The generation of a DFFS collection involves several stages as outlined in Figure 1.5. This process can be described as:

1. Collection of material from diverse geographic range.
2. Selection of representative subsets - ‘founder lines’.
3. Cross the founder lines to a rapid cycling, microspore responsive line, DHLS150.
4. Carried through microspore culture to produce doubled haploid lines with identical chromosomes.
5. Multiple the seed, archive and distribute the seed to researchers as required.

The founder lines are interesting to study as the genetic component of these lines is unaltered and as it would be in the wild. Many complications are associated with the use of the founder lines, including vast differences in development including differing germination rates, length of vegetative phase between lines and poor fecundity. In addition, the lines

are not genetically fixed, therefore each line is subject to recombination during meiosis, limiting the reproducibility between generations and replicability.

Using DH lines, in particular for RNAseq studies, has a multitude of benefits. This includes ease of read mapping, which is an important consideration as the C-genome *Brassica* species have a complex, highly duplicated genome and adding phase would add further complication to the analysis (Parkin et al., 2014; Wang et al., 2011a,b). With DH lines, individual genotypes can be replicated many times for increased statistical power. In addition, they should not exhibit allelic variation. The genomes of the DH lines are comprised of a mix of genetic material from the founder line used as a parent and the rapid cycler (DHSL150) that the founder line was crossed to. Thus, a further use of DH populations is the ability to map the introgression by comparing the sequence back to the parent in order to identify whether the genes of interest originated from the rapid cycling parent or the wild parent and to identify genes which underpin quantitative trait loci (QTL).

Generating the Cg-DFFS was carried out as part of DEFRA funded projects based at the University of Warwick. The Cg-DFFS collection has been previously screened for variation in multiple traits such as shoot Ca^{2+} and Mg^{2+} concentration (Broadley et al., 2008; White et al., 2010), seed oil content (Barker et al., 2007) and water use efficiency (Thompson, 2011, 2009) and was found to exhibit excellent genetic variability. This population has not previously been screened for salinity tolerance, therefore provides an untapped resource in the development of salt tolerant *B. oleracea* species.

1.6 Aims of the project

The overall aim of this project was to identify genes and pathways involved in the response to salt shock and other stress conditions on the transcriptome of C-genome *Brassica* species using transcriptomic technologies. In order to achieve this aim, the progression of the early stages of salt shock (0 - 36 hours) was thoroughly analysed through a high-resolution time-course analysis of the salt shock response in which several key genes and mechanisms were identified in the *B. oleracea* GD33DH broccoli line (Chapter 4). In order to achieve this, it was first necessary to design a *B. oleracea* specific microarray using newly available genome and transcriptomic data (Chapter 3), to include all gene models from the genome annotation and transcripts produced by *B. oleracea* GD33DH under various stress conditions. As the use of crop wild relatives provide a vast resource of genetic variation in response to stress conditions, an analysis of the salt shock response in wild *Brassica* species was conducted (Chapter 5) and potential genes which conferred tolerance in a tolerant species of C-genome *Brassica* were identified. The thesis concludes with an investigation into genome architecture in the salt shock and stress response in C-genome

Brassica species, highlighting the advantages of recent whole genome duplication events in the evolutionary history of *B. oleracea* (Chapter 6).

Chapter 2

Materials and Methods

2.1 Plant growth

The work described in Chapters 3 and 4 was carried out using the *B. oleracea* GD33DH line, a DH broccoli developed from the commercial line ‘Green Duke’, an F1 hybrid commercial cultivar bred by Sakata Seed Corporation (Uchaud, France). GD33DH has been used as a parent of several well studied mapping populations including the A12DH x GD33DH (AG) population (Bohuon et al., 1996, 1998; Broadley et al., 2008; Issa et al., 2013) and the Mar34 x GD33DH (MGDH) population (Skipper, 2010; Walley et al., 2011). In addition, a collection of C-genome *Brassica* species, described in Chapter 1 were used in phenotype screens subsequent RNAseq analysis in Chapter 5.

2.1.1 Growing *Brassica*

The appropriate C-genome *Brassica* seed was sourced from the Warwick Genetic Resources Unit, based at Wellesbourne, UK. An individual seed was sown on to F₁ compost (Levington Horticulture, UK), just under the soil surface (0.5 cm) in a P24 tray (Plankpak) in a chequered pattern, ensure each seedling had adequate space and watered generously. Trays containing seed were placed under natural light conditions either in glasshouses (Wellesbourne campus, UK) or in the Phytobiology facility (Warwick campus, UK) dependent on the experiment. Plants were watered as needed by Horticultural service technicians.

2.1.2 Experimental designs

Preliminary stress experiments

Experiments in Chapter 3 were conducted between October 2012 and January 2013. GD33DH seed was sourced from the Warwick Genetic Resources Unit, based at Welles-

Tray 1				Tray 2				Tray 3				Tray 4									
6hpt Salt B	6hpt Cont B	8hpt Salt B 8hpt Cont B	12hpt Salt A	12hpt Cont A				28hpt Salt B	28hpt Cont B	2hpt Salt A	2hpt Cont A	20hpt Salt A	0h Salt A	0h Cont A				26hpt Salt A	18hpt Salt A		
24hpt Salt A	24hpt Cont A		14hpt Salt A	14hpt Cont A				34hpt Salt A	34hpt Cont A			20hpt Cont A	6hpt Salt A	6hpt Cont A				28hpt Salt A	18hpt Cont A		
			30hpt Salt A	30hpt Cont A				32hpt Salt B	32hpt Cont B	32hpt Salt A	32hpt Cont A							16hpt Salt A	16hpt Cont A		
				22hpt Salt A	22hpt Cont A												8hpt Salt A	8hpt Cont A			
Tray 5				Tray 6				Tray 7				Tray 8									
		10hpt Salt B 10hpt Cont B	24hpt Salt C	24hpt Cont C	8hpt Salt C	8hpt Cont C	14hpt Salt C	14hpt Cont C	36hpt Salt A	36hpt Cont A	14hpt Salt B 14hpt Cont B	24hpt Salt B	24hpt Cont B	30hpt Salt B	30hpt Cont B			2hpt Salt B	2hpt Cont B		
30hpt Salt C	30hpt Cont C				34hpt Salt C	34hpt Cont C	12hpt Salt B	12hpt Cont B	34hpt Salt B	34hpt Cont B			0h Salt C	0h Cont C	4hpt Salt A	4hpt Cont A			22hpt Salt B	22hpt Cont B	
4hpt Salt C	4hpt Cont C		36hpt Salt B	36hpt Cont B	26hpt Salt B	26hpt Cont B	0h Salt D	0h Cont D	20hpt Salt B	20hpt Cont B						10hpt Salt A	10hpt Cont A				
Tray 9				Tray 10				Tray 11				Tray 12									
34hpt Salt D	34hpt Cont D	2hpt Salt C 2hpt Cont C	8hpt Salt D	8hpt Cont D	28hpt Salt D	28hpt Cont D	18hpt Salt C	16hpt Salt C	6hpt Salt C	6hpt Cont C	26hpt Salt C 26hpt Cont C	36hpt Salt C	36hpt Cont C					18hpt Salt B	18hpt Cont B		
22hpt Salt D	22hpt Cont D				10hpt Salt C	10hpt Cont C	26hpt Salt D	26hpt Cont D	30hpt Salt D	30hpt Cont D			36hpt Salt C	36hpt Cont C	22hpt Salt C	22hpt Cont C					
14hpt Salt D	14hpt Cont D								32hpt Salt C	32hpt Cont C				4hpt Salt B	4hpt Cont B	12hpt Salt C	12hpt Cont C			20hpt Salt C	20hpt Cont C
Tray 13				Tray 14																	
20hpt Salt D	20hpt Cont D	24hpt Salt D 24hpt Cont D			18hpt Salt D	18hpt Cont D			6hpt Salt D	6hpt Cont D	4hpt Salt D 4hpt Cont D										
32hpt Salt D	32hpt Cont D								10hpt Salt D	10hpt Cont D											
16hpt Salt D	16hpt Cont D				12hpt Salt D	12hpt Cont D			36hpt Salt D	36hpt Cont D								2hpt Salt D	2hpt Cont D		

Figure 2.1: Design of time-series experiment described in Chapter 4

GD33DH was glasshouse grown in the above design. For each time point (hours post treatment, hpt) and replicate (indicated by letters A - D). Treatment and control plants were paired and grown side-by-side to control for variation in glasshouse conditions.

Rep 1	Line 35 Saline	Line 35 Control	Line 42 Control	Line 42 Saline	Line 18 Saline	Line 18 Control	Line 01 Saline	Line 01 Control
	Line 33 Saline	Line 33 Control	Line 16 Control	Line 16 Saline	Line 12 Saline	Line 12 Control	Line 40 Saline	Line 40 Control
	Line 07 Control	Line 07 Saline	Line 28 Control	Line 28 Saline	Line 48 Control	Line 48 Saline	Line 24 Control	Line 24 Saline
	Line 23 Control	Line 23 Saline	Line 27 Saline	Line 27 Control	Line 06 Saline	Line 06 Control	Line 47 Control	Line 47 Saline
	Line 38 Saline	Line 38 Control	Line 21 Control	Line 21 Saline	Line 45 Saline	Line 45 Control	Line 04 Control	Line 04 Saline
	Line 19 Control	Line 19 Saline	Line 43 Control	Line 43 Saline	Line 36 Saline	Line 36 Control	Line 02 Saline	Line 02 Control
	Line 34 Control	Line 34 Saline	Line 13 Control	Line 13 Saline	Line 41 Control	Line 41 Saline	Line 17 Control	Line 17 Saline
	Line 14 Control	Line 14 Saline	Line 51 Saline	Line 51 Control	Line 31 Control	Line 31 Saline	Line 10 Control	Line 10 Saline
	Line 49 Control	Line 49 Saline	Line 29 Saline	Line 29 Control	Line 08 Control	Line 08 Saline	Line 25 Control	Line 25 Saline
	Line 37 Saline	Line 37 Control	Line 44 Control	Line 44 Saline	Line 20 Saline	Line 20 Control	Line 03 Control	Line 03 Saline
	Line 09 Saline	Line 09 Control	Line 30 Saline	Line 30 Control	Line 50 Control	Line 50 Saline	Line 26 Control	Line 26 Saline
	Line 52 Control	Line 52 Saline	Line 11 Saline	Line 11 Control	Line 15 Saline	Line 15 Control	Line 32 Control	Line 32 Saline
	Line 39 Control	Line 39 Saline	Line 05 Saline	Line 05 Control	Line 22 Saline	Line 22 Control	Line 46 Control	Line 46 Saline
Rep 2	Line 20 Saline	Line 20 Control	Line 06 Saline	Line 06 Control	Line 41 Saline	Line 41 Control	Line 35 Saline	Line 35 Control
	Line 16 Saline	Line 16 Control	Line 50 Control	Line 50 Saline	Line 02 Saline	Line 02 Control	Line 31 Saline	Line 31 Control
	Line 48 Control	Line 48 Saline	Line 14 Saline	Line 14 Control	Line 29 Control	Line 29 Saline	Line 13 Control	Line 13 Saline
	Line 19 Control	Line 19 Saline	Line 40 Saline	Line 40 Control	Line 34 Saline	Line 34 Control	Line 05 Saline	Line 05 Control
	Line 37 Control	Line 37 Saline	Line 08 Saline	Line 08 Control	Line 43 Control	Line 43 Saline	Line 22 Control	Line 22 Saline
	Line 28 Saline	Line 28 Control	Line 47 Saline	Line 47 Control	Line 26 Saline	Line 26 Control	Line 12 Control	Line 12 Saline
	Line 38 Saline	Line 38 Control	Line 09 Control	Line 09 Saline	Line 44 Control	Line 44 Saline	Line 23 Saline	Line 23 Control
	Line 52 Saline	Line 52 Control	Line 18 Control	Line 18 Saline	Line 04 Saline	Line 04 Control	Line 33 Saline	Line 33 Control
	Line 21 Control	Line 21 Saline	Line 42 Saline	Line 42 Control	Line 36 Saline	Line 36 Control	Line 07 Control	Line 07 Saline
	Line 11 Control	Line 11 Saline	Line 46 Control	Line 46 Saline	Line 25 Saline	Line 25 Control	Line 27 Saline	Line 27 Control
	Line 51 Control	Line 51 Saline	Line 32 Control	Line 32 Saline	Line 03 Saline	Line 03 Control	Line 17 Saline	Line 17 Control
	Line 49 Saline	Line 49 Control	Line 30 Saline	Line 30 Control	Line 15 Saline	Line 15 Control	Line 01 Saline	Line 01 Control
	Line 45 Saline	Line 45 Control	Line 39 Saline	Line 39 Control	Line 10 Saline	Line 10 Control	Line 24 Saline	Line 24 Control
Rep 3	Line 27 Control	Line 27 Saline	Line 01 Saline	Line 01 Control	Line 14 Saline	Line 14 Control	Line 40 Control	Line 40 Saline
	Line 46 Control	Line 46 Saline	Line 07 Saline	Line 07 Control	Line 33 Saline	Line 33 Control	Line 20 Control	Line 20 Saline
	Line 42 Control	Line 42 Saline	Line 16 Control	Line 16 Saline	Line 29 Control	Line 29 Saline	Line 03 Saline	Line 03 Control
	Line 06 Saline	Line 06 Control	Line 45 Saline	Line 45 Control	Line 32 Control	Line 32 Saline	Line 19 Saline	Line 19 Control
	Line 47 Control	Line 47 Saline	Line 08 Saline	Line 08 Control	Line 34 Control	Line 34 Saline	Line 21 Saline	Line 21 Control
	Line 25 Control	Line 25 Saline	Line 51 Control	Line 51 Saline	Line 12 Control	Line 12 Saline	Line 38 Control	Line 38 Saline
	Line 28 Saline	Line 28 Control	Line 41 Control	Line 41 Saline	Line 15 Control	Line 15 Saline	Line 02 Saline	Line 02 Control
	Line 31 Control	Line 31 Saline	Line 18 Saline	Line 18 Control	Line 44 Saline	Line 44 Control	Line 05 Control	Line 05 Saline
	Line 43 Saline	Line 43 Control	Line 17 Saline	Line 17 Control	Line 30 Control	Line 30 Saline	Line 04 Saline	Line 04 Control
	Line 39 Control	Line 39 Saline	Line 52 Control	Line 52 Saline	Line 13 Saline	Line 13 Control	Line 26 Saline	Line 26 Control
	Line 10 Saline	Line 10 Control	Line 36 Control	Line 36 Saline	Line 49 Control	Line 49 Saline	Line 23 Control	Line 23 Saline
	Line 35 Control	Line 35 Saline	Line 09 Saline	Line 09 Control	Line 48 Control	Line 48 Saline	Line 22 Control	Line 22 Saline
	Line 50 Saline	Line 50 Control	Line 11 Saline	Line 11 Control	Line 37 Saline	Line 37 Control	Line 24 Saline	Line 24 Control

Figure 2.2: Design of wild *Brassica* S1 experiment described in Chapter 5

Plants were glasshouse grown in the above design. Paired treatment and control plants were grown side-by-side to control for variation in glasshouse conditions.

Rep 1	C13012 Salt	C13012 Cont	C10121 Cont	C10121 Salt	C07069 Salt	C07069 Cont	TO1000 Salt	TO1000 Cont	C10128 Salt	C10128 Cont
	TO1000 Salt	TO1000 Cont	C13013 Cont	C13013 Salt	C10027 Salt	C10027 Cont	C07094 Salt	C07094 Cont	C10139 Salt	C10139 Cont
	C13013 Cont	C13013 Salt	C07019 Cont	C07019 Salt	C10128 Cont	C10128 Salt	C10025 Cont	C10025 Salt	Early Big Cont	Early Big Salt
	C10139 Cont	C10139 Salt	C10125 Salt	C10125 Cont	C10121 Salt	C10121 Cont	Missing	Missing	C13001 Cont	C13001 Salt
	C07007 Salt	C07007 Cont	C07079A Cont	C07079A Salt	Early Big Salt	Early Big Cont	C10027 Cont	C10027 Salt	C10125 Cont	C10125 Salt
	C07069 Cont	C07069 Salt	C07060 Cont	C07060 Salt	C07007 Salt	C07007 Cont	C10132 Salt	C10132 Cont	C07094 Salt	C07094 Cont
	C13001 Cont	C13001 Salt	DHSL150 Cont	DHSL150 Salt	C07019 Cont	C07019 Salt	C13012 Cont	C13012 Salt	C07060 Cont	C07060 Salt
	C10025 Cont	C10025 Salt	C10132 Salt	C10132 Cont	Missing	Missing	C07079A Cont	C07079A Salt	DHSL150 Cont	DHSL150 Salt
Rep 2	C07069 Salt	C07069 Cont	C10027 Salt	C10027 Cont	C10139 Salt	C10139 Cont	Early Big Salt	Early Big Cont	DHSL150 Cont	DHSL150 Salt
	C13001 Salt	C13001 Cont	C10132 Cont	C10132 Salt	C07019 Salt	C07019 Cont	C07094 Salt	C07094 Cont	C07069 Salt	C07069 Cont
	Early Big Cont	Early Big Salt	C13001 Salt	C13001 Cont	C07079A Cont	C07079A Salt	TO1000 Cont	TO1000 Salt	C07060 Salt	C07060 Cont
	Missing	Missing	C13013 Salt	C13013 Cont	C07007 Salt	C07007 Cont	DHSL150 Salt	DHSL150 Cont	C13012 Salt	C13012 Cont
	C13013 Cont	C13013 Salt	C07079A Salt	C07079A Cont	C10128 Cont	C10128 Salt	C10121 Cont	C10121 Salt	C07094 Cont	C07094 Salt
	C10132 Salt	C10132 Cont	C10128 Cont	C10128 Salt	TO1000 Salt	TO1000 Cont	C10125 Cont	C10125 Salt	Missing	Missing
	C10125 Cont	C10125 Salt	C07019 Cont	C07019 Salt	C13012 Cont	C13012 Salt	C10025 Salt	C10025 Cont	C10027 Salt	C10027 Cont
	C10025 Cont	C10025 Salt	C10139 Cont	C10139 Salt	C07060 Cont	C07060 Salt	C07007 Cont	C07007 Salt	C10121 Cont	C10121 Salt
Rep 3	C07019 Cont	C07019 Salt	C10132 Salt	C10132 Cont	TO1000 Salt	TO1000 Cont	C10121 Cont	C10121 Salt	DHSL150 Cont	DHSL150 Salt
	C07069 Cont	C07069 Salt	C10121 Salt	C10121 Cont	C13001 Salt	C13001 Cont	C10025 Cont	C10025 Salt	C07079A Cont	C07079A Salt
	C13012 Cont	C13012 Salt	Early Big Cont	Early Big Salt	C13013 Cont	C13013 Salt	C10132 Salt	C10132 Cont	C10139 Salt	C10139 Cont
	DHSL150 Salt	DHSL150 Cont	C13001 Salt	C13001 Cont	C10027 Cont	C10027 Salt	C07007 Salt	C07007 Cont	C10128 Salt	C10128 Cont
	Early Big Cont	Early Big Salt	C07094 Salt	C07094 Cont	C07019 Cont	C07019 Salt	C10139 Salt	C10139 Cont	Missing	Missing
	TO1000 Cont	TO1000 Salt	C10027 Cont	C10027 Salt	Missing	Missing	C07060 Cont	C07060 Salt	C10025 Cont	C10025 Salt
	C07079A Salt	C07079A Cont	C13012 Cont	C13012 Salt	C07007 Cont	C07007 Salt	C10125 Cont	C10125 Salt	C07094 Cont	C07094 Salt
	C10125 Salt	C10125 Cont	C10128 Salt	C10128 Cont	C07060 Salt	C07060 Cont	C13013 Salt	C13013 Cont	C07069 Cont	C07069 Salt

Figure 2.3: Design of wild *Brassica* DH experiment described in Chapter 5. Paired treatment and control plants were grown side-by-side to control for variation in glasshouse conditions.

bourne, UK. Plants were grown as described in Section 2.1.1 and were placed into a glasshouse compartment (GH15/43; Wellesbourne campus, UK) which was set to maintain temperatures of 15°C by day and 15°C at night using automatic vents at 17°C. Plants were grown with regular watering using tap water for 42 days (6 weeks).

Time-course experiment

Experiments in Chapter 4 were conducted between July 2013 and August 2013. GD33DH seed was sourced from the Warwick Genetic Resources Unit, based at Wellesbourne, UK. Plants were grown as described in Section 2.1.1 and were placed into a glasshouse compartment (GH15/43; Wellesbourne campus, UK) which was set to maintain temperatures of 15°C with automatic venting at 17°C, both day and night. Following growth for 28 days, plants were transplanted into 7cm x 7cm pots (Plantpak) and growth continued for a further 14 days until the plants were 42 days (6 weeks). There were four biological replicates per time point; replicates were arranged in loose blocks, within which sampling times were randomly allocated (as shown in Figure 2.1).

Cg-DFFS S1 screen

Experiments in Chapter 5 were conducted between August 2013 and October 2013 in the Grodome compartments located at the Phytobiology facility, University of Warwick campus, UK. Plants were grown as described in Section 2.1.1. The compartment was set to maintain day temperatures of 20°C and night temperatures of 8°C. At 21 days of growth, healthy plants were transplanted into larger 7cm x 7cm pots, arranged in experimental

designs indicated in Figure 2.2. Experiments were begun when plants were 42 day old, in which a salt shock (250 mM, as described below) was applied to the treatment plant and controls were watered as normal. Sampling of three biological replicates occurred at 24 hours post treatment (hpt), in which leaf #5 was detached from the plant, placed into a pre-labelled 50 ml Falcon tube (Fisher Scientific), snap frozen in liquid nitrogen and stored at -80°C.

Cg-DFFS DH screen

Experiments in Chapter 5 were conducted between January 2015 and March 2015 in the Grodome compartments located at the Phytobiology facility, University of Warwick campus, UK. Plants were grown as described in Section 2.1.1. The compartment was set to maintain day temperatures of 20°C and night temperatures of 8°C. At 21 days of growth, healthy plants were transplanted into larger 7cm x 7cm pots, arranged in experimental designs indicated in Figure 2.3. Experiments were begun when plants were 42 day old, in which a salt shock (250 mM, as described below) was applied to the treatment plant and controls were watered as normal. Sampling of four biological replicates occurred at 24 hpt, in which leaf #5 was detached from the plant, placed into a pre-labelled 50 ml Falcon tube (Fisher Scientific), snap frozen in liquid nitrogen and stored at -80°C.

2.1.3 Stress conditions

Salt shock

Experiments were carried out using concentrations of 250 mM (14.625 g of NaCl per 1 l of deionised water) and 500 mM. 200 ml of NaCl was applied per plant, ensuring plants were watered to excess to ensure an equal distribution of NaCl throughout the root system. Controls were watered with deionised water.

Cold stress

Plants were placed into a controlled environment growth chamber (Sanyo 970) in 14 hr light conditions ($120 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) at 2°C, 60% humidity and 350 ppm CO₂.

Infection with *Sclerotinia sclerotiorum*

Prior to infection, spores of *S. sclerotiorum* (isolate L6, grown by Dr. Andrew Taylor, Wellesbourne, UK) were harvested and suspended in sterile half strength potato dextrose broth (PDB) and filtered through glass wool cloth. The inoculum was prepared into fresh, sterile half strength PDB and the concentration adjusted to 100,000 spores/ml. The spore concentration was measured using a hemocytometer. Leaf #5 was detached

from GD33DH *B. oleracea* using a sharp blade and were placed on 0.8% w/v plant agar (Duchefa Biochemie) in propagator trays. Around 10 droplets (20 μ l) of the innoculum was placed onto each ‘infected’ leaf and similarly around 10 droplets (20 μ l) of half strength PDB were placed onto each ‘mock-infected’ leaf. The trays were covered and sealed and incubated in a control environment chamber (Sanyo 970) in 14hr light conditions (120 μ mol photons m⁻² s⁻¹) at 20°C, 90% humidity and 350ppm CO₂.

2.1.4 Harvesting and storage of leaf material

For the preliminary stress experiments (Chapter 3) and the time-series analysis (Chapter 4) sampling occurred by detaching a leaf (leaf #5, unless otherwise indicated) at the base of the petiole, placing the leaf into a 50 mL Falcon tube (Fisher Scientific) which was snap frozen in liquid nitrogen, freeze dried and stored at -20°C. For the RNAseq analysis described in Chapter 5, two leaf discs were taken from leaf #5, placed into a pre-labelled 2ml Eppendorf tube, flash frozen in liquid nitrogen and stored at -80°C until RNA extractions were carried out.

2.1.5 Phenotype measurements

Plant height

Photographs were taken of plants at the horizontal level at a distance of \sim 2.5 metres. A ruler was included in the photographs, to be used as a scale. Using ImageJ, a line was drawn across 1 cm of the ruler, and the scale was set to 1 cm (*Analyse > Set Scale...* in the *Known Distance* box enter ‘1’, click OK.). Draw a line from the base of the plant to the tip, measure the line (*Analyse > Measure*) and from the *Results* box, data from the *length* column were recorded (cm).

Leaf area

Leaf #5 was detached from the plant at the base of the petiole. Paired control and treated leaves were placed side by side, next to a ruler on a lighting stage and photographed at a height of 55 cm. Using ImageJ, a line was drawn across 1 cm of the ruler, and the scale was set to 1 cm (*Analyse > Set Scale...* in the *Known Distance* box enter ‘1’, click OK.). The image was made binary (*Process > Binary > Make Binary*), ensuring the leaf remained black, and the background white. The leaf mask was selected using the wand tool and measurements were taken (*Analyse > Measure*). From the *Results* box, data from the *Area* column were recorded (cm²).

Fresh/dry weight

All shoot material above the soil level was detached using a blade, and weighed using a balance to give fresh weight data (g). The material was placed into a prelabelled paper bag which was placed into a drying oven at 65°C for three days, until all moisture was removed. The dry material was subsequently weighed to give dry weight data (g).

Mineral content

0.1 - 0.5 g of *Brassica* leaf material was digested using 2 ml nitric acid (70%), in a microwave digestion system for 30 minutes. The prepared sample volume was brought up to 25 ml with deionised water. Prior to analysis, a 1 in 20 dilution using deionised water was made before the samples were run on the Agilent LC-ICP-MS in the Department of Chemistry, University of Warwick, UK. A calibration curve was generated using Sodium and Potassium Standards for ICP (Sigma-Aldrich). Na⁺ was measured at 589.592 nm and K⁺ measured at 766.49 nm. This work was carried out by Almustapha Lawal (University of Warwick, UK).

2.2 Laboratory methods

2.2.1 RNA extractions

Three glass beads were added to a pre-labelled 2ml Eppendorf tube prior to sampling. A sample of leaf was taken, placed into the Eppendorf tube, flash frozen in liquid nitrogen and kept on dry ice throughout the process. Eppendorf adaptor blocks were placed in a -80°C freezer for two hours, after which they were removed and the tubes containing the frozen leaf material were placed into the blocks. The frozen blocks were placed into the mixer mill MM400 (Retsch), to grind the leaf material into a fine powder (30 Hz, 1 minute per adaptor side, 2 minutes in total). If necessary, this process was repeated until a fine powder was achieved. The samples were kept on dry ice following milling. 1ml of TRIzol (Invitrogen) was added to each sample and vortexed for 30 seconds. Samples were incubated at room temperature for 5 minutes to allow for the dissociation of nucleoprotein complexes, then 200 μ l of chloroform was added. Samples were vortexed for 15 seconds and incubated for 3 minutes at room temperature. The samples were then centrifuged at 8000 g for 20 minutes at 4°C. The upper aqueous phase was removed, with care to ensure that the interphase was not disturbed and transferred to a fresh 1.5 ml Eppendorf tube. An equal volume of 70% ethanol made up with RNase free water was added mixed thoroughly by pipetting. 700 μ l of the sample, including any precipitate was transferred to an RNeasy purification column (QIAGEN). This was centrifuged at 8000 g for 15 s at room temperature and the flow-through discarded. An on-column DNase digestion

(RNase-Free DNase set, Qiagen) followed by RNA cleanup was carried out according to manufacturers instructions (QIAGEN RNeasy Mini Kit, Part 2). The RNA was eluted in 30 μ l of RNase-free water (Qiagen), 3 μ l of the elute was aliquoted for quantification and quality control. The eluted RNA was stored at -80°C. Concentration and purity was measured using a Nanodrop ND-1000 spectrophotometer (Thermo scientific) using a 1.5 μ l sample. The quality of RNA was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies) with an RNA 6000 Nano chip (part number 5067-1511) according to manufacturers instructions.

2.2.2 qPCR

cDNA synthesis was performed using SuperScript II Reverse Transcriptase (Invitrogen). 1 μ l 50 mM oligo(dT)₁₈ and 1 μ l 10 mM dNTPs were added to the DNase treated RNA sample, before being incubated at 65°C for 5 minutes to anneal oligos to RNA. 4 μ l First Strand Buffer, 2 μ l dithiothreitol (0.1M), 1 μ l RNase OUT Recombinant Ribonuclease Inhibitor (Invitrogen) and 1 μ l SuperScript II Reverse Transcriptase (Invitrogen) was added to each sample before being incubated at 42°C for 50 minutes, followed 70°C for 15 minutes to inactivate the enzyme. If necessary, cDNA samples were stored at -20°C.

For qPCR analysis, primers specific to the target gene were designed to amplify 50-150bp of the coding sequence using NCBI primer blast (Table 2.1). Primers designed for amplification of the *PUX1* transcript (Bo7g084420.1) were used as a sample control for each sample. cDNA samples were diluted by 10 before qPCR analysis (initial concentration of 50 ng/ μ l). 5 ng of cDNA was mixed with 5 μ l of SsoAdvanced SYBR Green Supermix (Bio-Rad) and primers specific for the gene target (200 nM), to a total volume of 10 μ l. Each reaction was performed in triplicate as technical replicates. In addition, for every primer mix a non-template control was included to ensure the reaction mix was not contaminated and a standard curve was included by mixing equal volumes of every sample in the reaction, before serial dilution by 5 multiple times. qPCR reaction cycle was performed on a CFX384 Touch Real-Time PCR Detection platform (Bio-Rad) in 384-well white skirted BioRad qPCR plates. A 2-step PCR reaction was used, with a pre-cycle 95°C for 3 minutes, followed by 45 cycles of 95°C for 10 seconds, 55°C for 30 seconds. Fluorescence of each well was recorded after each cycle. A post-reaction melt-curve was performed by heating the sample to 95°C for 10 seconds, then performing a temperature gradient increase of 65°C to 95°C at 5 second increments. Fluorescence was measured after each temperature increase. A single melt-curve peak was confirmed visually.

Target Gene	Direction	Oligo sequence (5'-3')
PR1	Forward	CAGCCTTCGCTCAAAGCTAC
PR1	Reverse	GAAAAGTCGGCGCTACTCCA
COR15a	Forward	AGGAAACGAAGCTGGGAACA
COR15a	Reverse	TTTTGTGGCGTCCTTAGCCT
JAZ1	Forward	GCTTCTCGCTGACGTGTAGT
JAZ1	Reverse	GCTTACGTGACATGCCGTTG
RD26	Forward	TTGCCTTGAAGACCACAGCA
RD26	Reverse	AGCCCATTCGAAATTCCTCGT
PUX1	Forward	TGACCCAACGCTACTGACATC
PUX1	Reverse	GCATACCAGCAGCGACCTTA

Table 2.1: qPCR primer sequences

2.2.3 Microarrays

Custom designed Agilent 4x180K microarrays were used in Chapter 4, using kits and instructions supplied by the manufacturers (Agilent Technologies).

Labelling RNA

Quality of a selection of RNA samples was tested prior to labelling using an Agilent 2100 Bioanalyzer (Agilent Technologies) with an RNA 6000 Nano chip (part number 5067-1511) according to manufacturers instructions, to ensure quality was sufficient. Agilent RNA Spike In Kit for Two color v4.0 (part number 5188-5279) was added to the mRNA prior to the labelling reactions. These are control targets comprised of a set of ten *in-vitro* synthesized polyadenylated transcripts which were derived from the adenovirus E1A gene at known concentrations spanning a 200-fold dynamic range used for monitoring performance and quality of the labelling reaction.

100ng of total RNA plus spike mix was added with an Oligo dT-Promoter Primer to amplify Poly A⁺ mRNA samples from stress treated and control *B. oleracea* GD33DH. Labelling with the Agilent LowInput QuickAmp Labeling Kit Two-Color (part number 5190-2306) was carried out on the cDNA as per the recommended protocol - Two-Color Microarray-Based Gene Expression Analysis (Low Input Quick Amp Labeling) Protocol, version 6.6 (Sept. 2012). Poly A⁺ RNA from each sample was labelled in the presence of cyanine 3-CTP or cyanine 5-CTP (provided in the kit, part number 5190-2306) in separate labelling reactions, and stored on ice. cRNA concentration and Cy3/Cy5 dye concentration of 1.5 μ l of sample was measured using a Nanodrop ND-1000 spectrophotometer (Thermo scientific). The specific activity was calculated as per the protocol and a sample was passed onto the next stage should 260/280 ratio >1.8, cRNA yield >0.825 μ g and specific activity >6 (pmol Cy3 (or Cy5) per μ g cRNA). Multiple labelling reactions were performed and the labelled cRNA targets were stored at -80°C prior to use.

Slide-Position-Array#	Cy3	Cy5	Slide-Position-Array#	Cy3	Cy5
1-1-1	00SaltA	06ContB	16-1-61	14ContA	14SaltB
1-2-2	12SaltA	14ContA	16-2-62	36ContC	00SaltA
1-3-3	34ContB	36ContB	16-3-63	36ContA	00SaltB
1-4-4	18ContC	20SaltA	16-4-64	12SaltC	14SaltC
2-1-5	10SaltB	12SaltB	17-1-65	34SaltA	02SaltB
2-2-6	30ContB	32SaltB	17-2-66	30ContA	30SaltB
2-3-7	18ContA	26SaltB	17-3-67	20SaltC	22SaltC
2-4-8	36ContB	04ContC	17-4-68	28ContA	30SaltA
3-1-9	24ContB	24SaltC	18-1-69	22SaltC	28SaltC
3-2-10	00ContB	00SaltC	18-2-70	30SaltC	36SaltC
3-3-11	18ContB	20ContB	18-3-71	06ContC	12ContC
3-4-12	04ContC	06ContC	18-4-72	08SaltA	10SaltA
4-1-13	26SaltC	28ContA	19-1-73	32SaltC	34ContC
4-2-14	30ContC	36ContC	19-2-74	12ContB	18ContB
4-3-15	16ContB	16SaltC	19-3-75	22ContB	24SaltB
4-4-16	08ContB	08SaltC	19-4-76	02SaltA	08SaltA
5-1-17	14ContB	16SaltB	20-1-77	32SaltA	34SaltA
5-2-18	04ContB	10ContB	20-2-78	28ContB	34ContB
5-3-19	00SaltB	00ContC	20-3-79	18SaltA	24SaltA
5-4-20	06SaltC	12SaltC	20-4-80	16SaltC	18ContA
6-1-21	18SaltB	20SaltB	21-1-81	16SaltA	18SaltA
6-2-22	20SaltB	26SaltB	21-2-82	00SaltA	02SaltA
6-3-23	20ContA	22SaltA	21-3-83	24SaltB	24ContC
6-4-24	36ContC	00ContA	21-4-84	00SaltC	02ContC
7-1-25	36SaltC	00SaltA	22-1-85	06ContA	06SaltB
7-2-26	14SaltA	14ContB	22-2-86	14ContC	20ContC
7-3-27	10SaltA	16SaltA	22-3-87	16SaltB	16ContC
7-4-28	16SaltC	18ContC	22-4-88	32ContA	34ContA
8-1-29	18ContA	24ContA	23-1-89	10ContC	12SaltA
8-2-30	22ContC	28ContC	23-2-90	02ContA	08ContA
8-3-31	02ContB	04ContB	23-3-91	10ContB	12ContB
8-4-32	08SaltC	10ContC	23-4-92	00ContA	02ContA
9-1-33	16ContC	18SaltC	24-1-93	04ContA	06SaltA
9-2-34	32ContC	34SaltC	24-2-94	18SaltC	20ContA
9-3-35	06ContB	16SaltC	24-3-95	08ContC	10SaltC
9-4-36	30SaltA	30ContB	24-4-96	20ContC	22ContC
10-1-37	06SaltB	08ContB	25-1-97	34ContC	36SaltA
10-2-38	14SaltB	16ContB	25-2-98	30SaltB	32ContB
10-3-39	32ContB	32SaltC	25-3-99	26SaltA	32SaltA
10-4-40	22SaltA	22ContB	25-4-100	02SaltC	04ContA
11-1-41	28ContC	30ContC	26-1-101	04SaltA	06ContA
11-2-42	24ContA	26ContA	26-2-102	26ContB	28ContB
11-3-43	24SaltA	26SaltA	26-3-103	24SaltC	26ContC
11-4-44	26ContA	32ContA	26-4-104	02ContC	04SaltA
12-1-45	00ContC	02SaltC	27-1-105	12SaltB	18SaltB
12-2-46	20ContB	26ContB	27-2-106	16ContA	18ContA
12-3-47	12ContC	14ContC	27-3-107	10ContA	16ContA
12-4-48	20SaltA	22ContA	27-4-108	36SaltB	04SaltC
13-1-49	02SaltB	04SaltB	28-1-109	28SaltB	34SaltB
13-2-50	08SaltB	08ContC	28-2-110	12ContA	14SaltA
13-3-51	34ContA	02ContB	28-3-111	34SaltB	36SaltB
13-4-52	26SaltB	28SaltB	28-4-112	22ContA	22SaltB
14-1-53	24ContC	26SaltC	29-1-113	04SaltB	10SaltB
14-2-54	14SaltC	20SaltC	29-2-114	28SaltC	30SaltC
14-3-55	34SaltC	36ContA	29-3-115	06SaltA	06ContB
14-4-56	22SaltB	24ContB	29-4-116	36SaltA	00ContB
15-1-57	26SaltB	36ContC	30-1-117	28SaltA	30ContA
15-2-58	04SaltC	06SaltC	30-2-118	26ContC	28SaltA
15-3-59	06ContB	08SaltB	30-3-119	08ContA	10ContA
15-4-60	32SaltB	32ContC	30-4-120	10SaltC	12ContA

Table 2.2: Microarray hybridization plan

‘Slide’ refers to an entire 4x180 microarray platform, ‘position’ to the array within this platform and ‘array #’ is the number assigned to each individual array. Cy3-/Cy5-labelled samples described by sampling time (hpt), treatment followed by replicate e.g. ‘24ContB’ refers to a control sample, replicate B sampled at 24 hpt.

Microarray hybridization and scanning

The labelled cRNA targets were hybridized to SurePrint G3 Custom GE 4x180K Microarrays (G4862A) using the Agilent Gene Expression Hybridization Kit (part number 5188-5242) containing 1.65 μ g of Cy3-labelled and 1.65 μ g of Cy5-labelled cRNA per hybridization, as per the hybridization design (Table 2.2). The hybridization reactions were performed at 65°C for 17 hours in Agilent SureHyb ovens (G2545A), after which the arrays were removed and washed according to protocol.

Fluorescence (at 532 and 633 nm for Cy3 and Cy5, respectively), was measured using an Agilent microarray scanner (G2565CA). Quantification of features following scan was carried out using Agilent Feature Extraction Software (v10.7.3) with the GE2_107_Sep09 protocol using a custom grid provided by Agilent alongside the array design (AMADID 068323).

2.2.4 RNAseq: library preparation and sequencing

Prior to library preparation, mRNA quality was determined using a Nanodrop ND-1000 spectrophotometer (Thermo scientific) and an Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Kit (part number 5067-1513). mRNA with 260/280 and 260/230 ratios of <1.8 and a clean bioanalyzer trace were sent for sequencing.

The RNAseq library preparation and sequencing was outsourced to the Genome Centre at Barts and the London School of Medicine and Dentistry (London, UK) in October 2013 (as discussed in Chapter 3). The library prep for RNAseq described in Chapter 5 was carried out by the Genomics Facility at the University of Warwick and sequencing was carried out at the Wellcome Trust Centre for Human Genetics (Oxford, UK) in August 2015.

For both RNAseq experiments, library preparation was carried out externally as described, using Illumina TruSeq RNA library prep kit (v2). The sequencing was carried out using an Illumina HiSeq 2000 (Illumina, 2011). This resulted in 2x100bp paired-end reads, which were received back in FASTQ file format.

2.3 Bioinformatics methods

2.3.1 MAANOVA

An in-house adapted package MAANOVA (MicroArray ANalysis Of VAriance), implemented in R (McHattie, 2011), was used for the statistical analysis of the gene expression data generated as a result of the time-series experiment (Chapter 4). This program can handle time-series data and is used for assessing data quality, applying data transforma-

tions, fitting ANOVA models to estimate relative gene expression levels and carrying out F -tests for differential expression analysis. The co-ordinating script for this package can be found in Appendix A. Further details are supplied in Chapter 4.

2.3.2 Gaussian Process Two Sample (GP2S)

A locally adapted version of the Time-Local Gaussian Process Two Sample (GP2S) test (Stegle et al., 2010) was applied to the expression data by Dr. Christopher Penfold. The Gaussian Process determines differential expression by calculating a Bayes factor between two models. One model assumes that the expression profiles (Treatment/Control) are drawn from an identical, shared distribution and the alternative model assumes each expression profile comes from two independent distributions.

The timing of differential expression can be identified using a mixture model, switching between the two hypotheses, corresponding either to the shared model (can be represented by a single Gaussian Process) or the independent model (represented by two Gaussian processes) as a function of time.

The output of this analysis was a ranked Bayes Factor score per-probe from most to least like to be differential expressed and a Z-indicator of Gaussian noise score for each differentially expressed probe indicating the divergence of expression profiles at each time point.

2.3.3 Multiple Dataset Integration (MDI)

Clustering of expression profiles produced as part of the time-series analysis (described in Chapter 4) was carried out using Multiple Dataset Integration (MDI) (Kirk et al., 2012; Mason et al., 2016). MDI aims to share clustering correlations across many related datasets, such as multiple time-series experiments. The clustering algorithm uses a flexible Bayesian mixture modelling approach, letting the natural number of clusters that exist across the multiple datasets be determined without relying on overly strong modelling assumptions. This allows for clustering of expression profiles with very specific characteristics.

The output is a Markov Chain Monte Carlo (MCMC) analysis from which a Posterior Similarity Matrix (PSM) for each dataset and finally a consensus (mean) PSM is generated. The consensus PSM can be used to partition the expression profiles into cluster groups, in which expression profiles showing co-expression across the multiple datasets could be found in the same cluster. To implement MDI, the following command was used:

```
$ mdi++ GP controldata.csv saltdata.csv -c 200 -t 10 > output.csv
```

in which GP indicates that data should be loaded as Gaussian Process data, `-c` is the

maximum number of clusters to be filled, `-t` is to thin the output to one sample every 10 samples to decrease the size of the output.

2.3.4 Causal Structural Inference (CSI)

The Causal Structure Identification (CSI) GUI (Penfold et al., 2015; Penfold and Wild, 2011), implemented in MATLAB, was used to infer network topology using high-resolution, genome-wide time-series data described in Chapter 4.

Each gene/expression profile was set to be both a target and putative transcription factor to determine an unconstrained network. A non-hierarchical algorithm was selected, and the inference method used was Expectation Maximisation (EM), with default parameters.

The output of this analysis was a *.sif* file containing, for each gene and potential regulator, a thresholdable set of the marginal probability that the TF expression profile has an influence on the target gene expression profile. A marginal probability of 0.1 was used producing a dense, but not insignificant regulatory network. This output could be visualised in Cytoscape (v3.1.0) to produce a gene regulatory network model, alongside marginal probabilities to show the strength of the interaction.

2.3.5 RNAseq: data analysis

Preprocessing RNAseq reads

Before RNAseq analysis took place, an essential quality control step took place on the raw data to ensure it was of adequate quality, ensuring the reliability of results. Quality was checked using FastQC, a quality control tool for high throughput sequence data. This program was used to determine quality (Phred score) at each base pair throughout all reads, highlighting possible machine sequencing errors, poor quality reads and over-represented sequences indicating the presence of adaptors and primers.

Phred scores were used to assess data quality, a phred score $<Q20$ indicated poor quality data, $Q20 - Q30$ indicated data of intermediate quality and finally a phred score $>Q30$ indicated high quality data. Should data dip below $Q30$ for a large proportion of the reads, preprocessing of the reads was considered necessary ensuring only high quality reads were passed to the aligner or assembler. Where necessary, this preprocessing was carried out using Trimmomatic, which accepts paired-end reads.

```
$ java -jar ./trimmomatic.jar PE -threads 4 -phred33 -trimlog trimlog
File.txt -basein R1.fastq R2.fastq SLIDINGWINDOW:4:20 ILLUMINACLIP:ov
errepresented-adaptor-seqs.fa
```

Genome-guided RNAseq analysis

The inner mate pair distance is a metric required by TopHat and represents the distance between the two reads. This was calculated by carrying out a Bowtie run on the data with the minimum and maximum insert size set to 0 and 500 respectively:

```
$ bowtie-build Boleracea.v1.cds/genome.fasta Boleacea
$ bowtie Boleracea -1 R1.reads.fastq -2 R2.reads.fastq -I 0 -X 500 -S
```

where `-I` and `-X` refer to the minimum and maximum insert size and `-S` indicates the output should be in sam format. The insert size was calculated using Picard tools:

```
$ java -Xmx2g -jar picard-tools/CollectInsertSizeMetrics.jar INPUT=bowt
iealignment.sam HISTOGRAM_FILE=InsertSizeMetricsHist.pdf OUTPUT=InsertS
izeMetrics.txt
```

The inner mate distance was estimated by:

$$\text{Mean insert size} - 2 \times \text{read length} = \text{Inner mate distance}$$

TopHat2 (v2.2.1) was used to align the reads to the reference *B. oleracea* (TO1000) genome (Trapnell et al., 2012a) using the following parameters:

```
$ tophat Boleracea.v1.genome.fasta sample1-1.fastq.gzv sample1-2.fastq.gz -
r 74 -i 50 -I 50000 -p 8 --no-mixed --transcriptome-index Boleracea.v1.
cds.fasta
```

in which compressed fastq read files were aligned to the *B. oleracea* transcriptome with the following parameters: `-r` is the inner mate distance, as calculated above, `-i` minimum intron size, `-I` maximum intron size, `-p` the number of threads, finally the `--no-mixed` option prevents the reporting of reads where only one of the read pairs has aligned. The output was a .bam file containing the sequence alignment data in binary format which can be used in downstream steps.

Cufflinks (v2.2.1) was used to assemble the transcriptome containing novel transcripts, producing a gtf/gff (Gene Transfer Format/General Feature Format) file. Cuffmerge (v2.2.1), followed by the gffread function supplied with the Cufflinks software was used to merge multiple gtf files and produce a multifasta file of all transcript sequences, as per the following:

```
$ samtools sort bamfile1.bam sortedbamfile1
```

which sorts the alignments by the leftmost coordinate. The transcriptome is assembled using Cufflinks and multiple gtf file were merged using Cuffmerge:

```
$ cufflinks -g Boleracea.v1.genes.gff3 -o sortedbamfile1.bam
```

```
$ cuffmerge -o out -g Boleracea.v1.genes.gff3 -s Boleracea.v1.genome.  
fasta gtf-to-merge.txt
```

in which the *gtf-to-merge.txt* file contains a list of all the gtf/gff files that are to be merged in this step, all to be found in the same directory.

```
$ gffread -w output.fasta -g Boleracea.v1.genome.fasta merged.gtf
```

where *-w* is the name of the file to write the output to, *-g* is the genome from which to extract the sequence data. This produced a multi-fasta file of all features present in the merged.gtf file, as determined from the *B. oleracea* genome sequence.

***De-novo* assembly of RNAseq reads**

De-novo of RNAseq reads was carried out using the Trinity software (Haas et al., 2013) using iPlant Collaborative cloud computing resources (Goff, 2011). Firstly, all pre-processed read files were concatenated in identical order:

```
$ cat S1_R1.fastq...Sx_R1.fastq > All.R1.fastq  
$ cat S1_R2.fastq...Sx_R2.fastq > All.R2.fastq
```

Read files were then normalised by k-mer coverage to reduce computational time:

```
$ Trinity_normalize_by_k-mer_coverage --left=./All.R1.fastq --right=./  
All.R2.fastq --seqType=fq --max_cov=30 --kmer_size=25 --max_pct_stdev=100
```

where *--seqType* indicates the type of input file (fastq), *--max_cov* is the targetted maximum coverage for reads, *--kmer_size* is the kmer size and *--max_pct_stdev* is the maximum pct of mean for standard deviation of kmer coverage across the read. The output of this was a normalized version of all the reads to be written, reducing the memory and time taken to run the Trinity software in subsequent steps:


```
$ TRINITY_HOME/Trinity --seqType=fq --left=./norm_R1.fastq --right=./  
norm_R2.fastq --outputAssemblyFasta=Trinity.fasta
```

The output of this was a *de-novo* assembly of all reads in a multi-fasta file format.

Removing redundant transcripts from multi-fasta file

CD-HIT est (Li and Godzik, 2006) was used to cluster together highly homologous sequences to reduce the size of large multi-fasta files. CD-HIT est was implemented using the following options:

```
$ cd-hit-est -i input.fasta -o output.fa -c 0.95
```

where `-c` is the sequence identity threshold of similarity at which to collapse similar sequences into a cluster.

Counting RNAseq reads

Co-ordinate sorted .bam files were passed to HTSeq-count, which counts the number of reads aligning to each feature, here each gene model was used as the feature.

```
$ samtools sort bamfile1.bam sortedbamfile1
```

```
$ htseq-count sortedbam1.bam Boleracea.v1.genes.gff3 -f bam -r pos -t  
gene_id -m union > counts.txt
```

where `-f` indicates the file type, `-r` is used to determine how paired-end data is sorted, in this case by co-ordinate (as above), `-t` is the feature type to be used for counting, in this case `gene_id` was the preferred option and finally `-m` is the method to be used to handle reads which overlap more than one feature, union being the preferred option.

Identifying differentially expressed genes using count data

DESeq2 (Love et al., 2014) was used for differential expression analysis of replicated RNAseq count data. DESeq2 models read counts as following a negative binomial distribution and uses Empirical Bayes shrinkage for dispersion estimation and fold change estimation. Finally, a Wald test produces a *p*-value by comparing the beta estimate divided by its estimated standard error to a standard normal distribution, the output of which is adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). The co-ordinating script for this can be found in Appendix B.

2.3.6 Creating and querying a BLAST database

Blast database (version 2.2.28+) was created using the following:

```
$ makeblastdb -in Boleracea.v1.cds.fasta -db_type nucl -out Boleracea.v1
```

This resulted in the generation of files with the prefix ‘Boleracea.v1’ .fai, .nhr, .nin, .nsd, .nsq in the working directory following construction of the BLAST database. To query a standalone blast database (version 2.2.28+) with the query sequence(s) in a multi-fasta file:

```
$ blastn -query query.fasta -db Boleracea.v1 -evaluate 1e-20 -outfmt 6 -num_threads  
6 > blast_out.txt
```

The E value was set at $1e^{-20}$, however this could be lowered if a less stringent alignment was required. If only the top hit was required, `-max_target_seqs 1` was added. The output format was a tabulated text file, should an xml file be required, substitute `outfmt -6` for `outfmt -5`.

2.3.7 GO term enrichment analysis

During statistical analysis of GO terms, using BiNGO (Maere et al., 2005b), hypergeometric tests are used to identify whether a list of genes contain a larger proportion of members assigned with specific GO terms than the expected number, given the abundance of that GO term throughout the genome (background set). An adjusted *p*-value was output using Benjamini and Hochberg correction (Benjamini and Hochberg, 1995). When $p < 0.05$ the GO term was considered to be over-represented. BiNGO is capable of producing graphical representation of over-represented GO terms in the form of hierarchical graphs and text for further analysis.

2.3.8 Phylogenetic tree construction

SNP data, arranged in multifasta file format, was submitted to the Clustall Omega online portal (EMBL European Bioinformatics Institute, UK) using the Neighbour joining method without distance correction (McWilliam et al., 2013).

2.3.9 Determining diurnal expression

Transcript expression profiles that exhibited diurnal expression were identified using JTK_CYCLE with default parameters (Hughes et al., 2010). Expression profiles were considered diurnal if the adjusted *p*-value > 0.05.

Chapter 3

Development of transcriptomic resources in *Brassica oleracea*

3.1 Chapter overview

Since plants are unable to escape from their environment, they must have the ability to rapidly respond to a variety of unstable, potentially stressful conditions. This involves reprogramming a large proportion of the transcriptome, to enable the plant to redirect its energies into stress tolerance and survival rather than continued growth. Extensive changes have been shown to occur in plants in response to different types of stress conditions such as pathogen infection, drought and heat (Shinozaki et al., 2003; Wang et al., 2003; Windram et al., 2012).

Genome-wide changes in gene expression can be quantified using microarrays, a method of analysis that uses hybridisation of fluorescently labelled mRNA to complementary probes, to measure gene expression levels within a sample. Community resources previously available for *B. oleracea* included an Agilent microarray containing 44k probes designed from *B. napus* in 2008, a 135K unigene Affymetrix GeneChip of the A&C genome (Love et al., 2010) and a 95K community Brassica array containing EST sequences from *B. napus*, *B. rapa*, and *B. oleracea* (Trick et al., 2009). Since the release of the *B. oleracea* TO1000 genome (Parkin et al., 2014) and the availability of RNAseq data generated from *B. oleracea* and other C-genome *Brassica* species, these array designs were considered outdated.

For this thesis, the transcriptome sequence was updated and extended using new genomic information. RNAseq data (including some stress related data described in this chapter) and the *B. oleracea* TO1000 genome sequence have been mined to identify an inclusive list of transcripts for *B. oleracea*. In addition, to widen the application of the microarray to include *Brassica* C-genome species, including wild *B. oleracea*, RNAseq data from the CgDFFS was also mined to identify additional sequence information. This

knowledge has been used in the design of a new microarray which was used for subsequent experiments.

The aim of this chapter was to generate a new microarray resource for *B. oleracea* by carrying out the following objectives:

1. Investigate stress responses in *B. oleracea* GD33DH to determine suitable experimental conditions.
2. Using RNAseq, identify stress related transcripts for *B. oleracea* GD33DH.
3. Design a new microarray for the *B. oleracea* transcriptome using a range of genomic information.
4. Annotate the *B. oleracea* TO1000 genome with GO terms.

3.2 Stress response pathways are activated in *B. oleracea* GD33DH

Abiotic and biotic stress experiments were carried out using GD33DH, to develop suitable methodology for future stress treatments and to obtain stressed leaf material for RNA sequencing. Plants were subjected to various stress conditions including cold, salt shock and infection with *Sclerotinia sclerotiorum*, a necrotrophic pathogen. Three biological replicates for the treatment were harvested alongside three controls for each stress condition. Each biological replicate was whole leaf 5 from an individual GD33DH plant.

Plants subject to cold stress were placed into a controlled environment chamber (Sanyo) at 2°C for 24 h, after which leaves were harvested. The high salinity treatment was applied as a ‘salt shock’ in which a high concentration of saline solution (250 mM and 500 mM) was applied to the soil in excess whilst control plants were watered as normal. These values were chosen based on literature search of a moderate to severe salt stress in *Brassica* species (Mittal et al., 2012; Sharma et al., 2015). Finally, infection with *S. sclerotiorum* was carried out as a detached leaf assay (described in Chapter 2). Leaves were kept in a sealed tray in a growth cabinet at high humidity and harvested after 48 h.

The onset of stress was determined by the differential expression of key stress marker genes. Changes in gene expression occur at the onset of stress whilst a stress phenotype would take days to weeks to appear by which time gene expression would be at the acclimation stage rather than initial stress response. As such, no quantitative phenotype measurements were taken during this stress experiment, however the appearance of the treated plants was noted both at sampling time and, in the case of salt shock, in equivalent plants left for longer periods after the treatment. As shown in Fig. 3.1a, plant height was

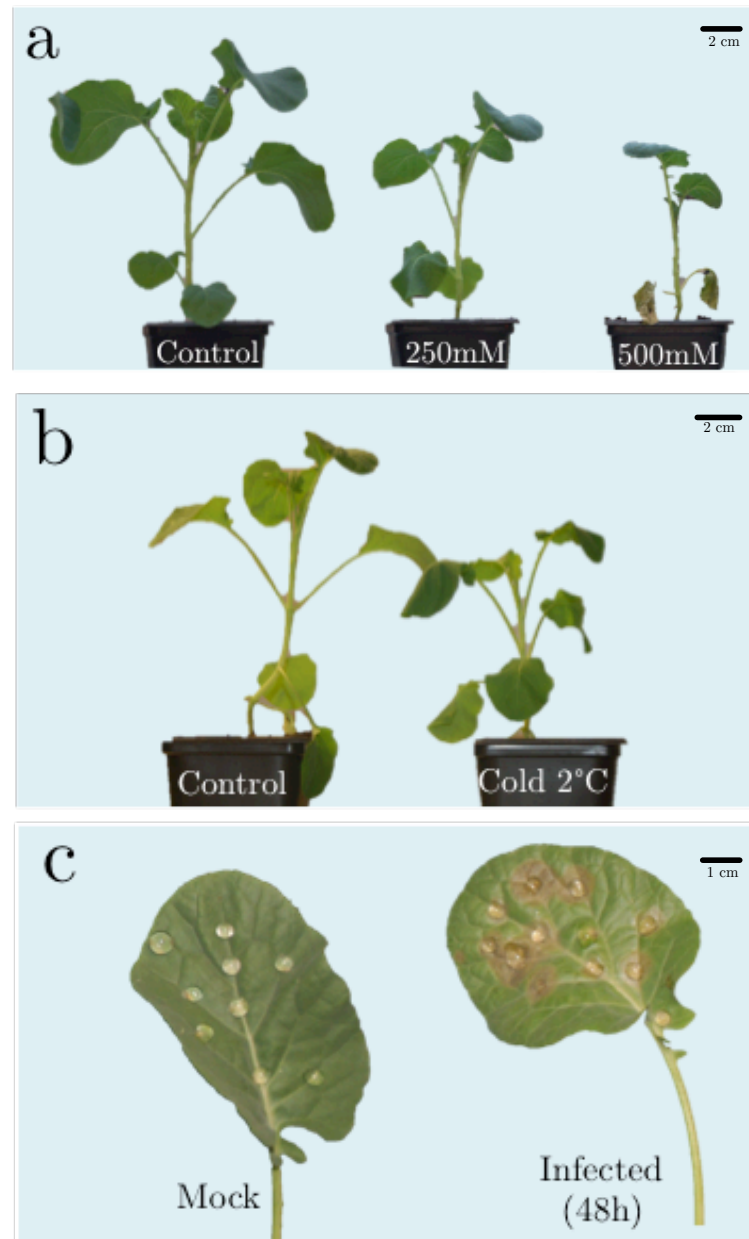
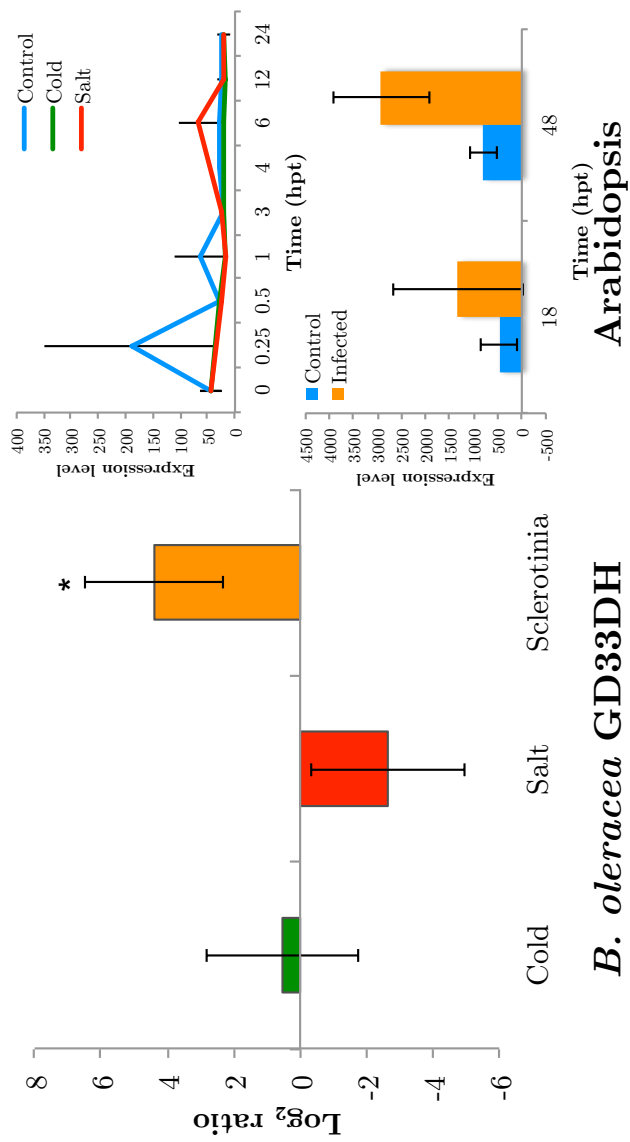


Figure 3.1: Investigation of the effects of different stress treatments on *B. oleracea* GD33DH

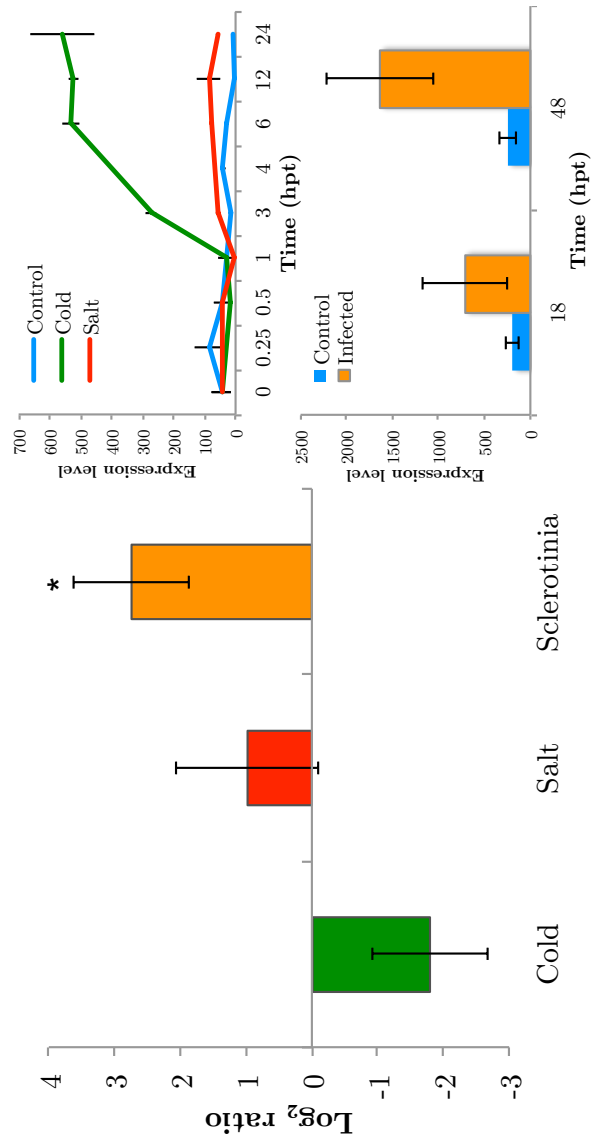
Typical phenotype of *B. oleracea* GD33DH following (a) salt shock six days after treatment (control on left, 250 mM and 500 mM NaCl), (b) cold stress, plants held at 2°C for 24 h (cold stress), control on left, treatment on right and (c) detached leaf 5 with mock inoculum and *S. sclerotiorum* infected droplets of inoculum leading to the formation of necrotrophic lesions on the leaf, 48 h post infection. Scale is indicated by the bar.

severely affected by the high salinity treatment 6 days after treatment. The cold stress (Fig. 3.1b), resulted in wilting after 24 h at 2°C. The effect of height seen in Fig. 3.1b was likely due to natural variation in height of the individual plants, before the treatment was applied. Successful infection with *S. sclerotiorum* was clearly seen by the formation of large necrotic lesions on the leaf (Fig. 3.1c).

In order to ensure that the treated plants were experiencing stress conditions, expression of key marker genes were measured using qPCR. The null hypothesis of the experiment was ‘there is no difference in expression of stress response genes in control versus treatment’. Each gene was measured in three biological replicates, and three technical replicates were included on each qPCR plate. *PUX1* was used as a housekeeping gene for normalisation during the analysis. The genes tested were *PR1*, *JAZ1*, *COR15a* and *RD26* as these have been reported to show stress enhanced expression in Arabidopsis (Fujita et al., 2004; Thines et al., 2007; Wang and Hua, 2009; Wu et al., 2009).



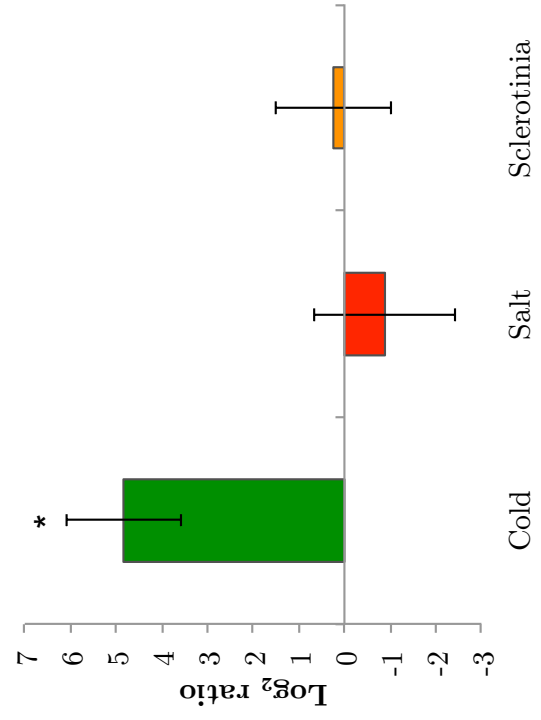
(a) *PR1*



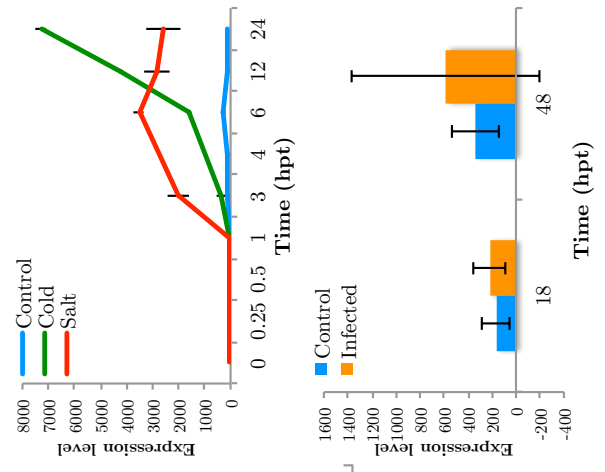
B. oleracea GD33DH

Arabidopsis

(b) *JAZ1*



B. oleracea GD33DH



Arabidopsis

(c) *COR15A*

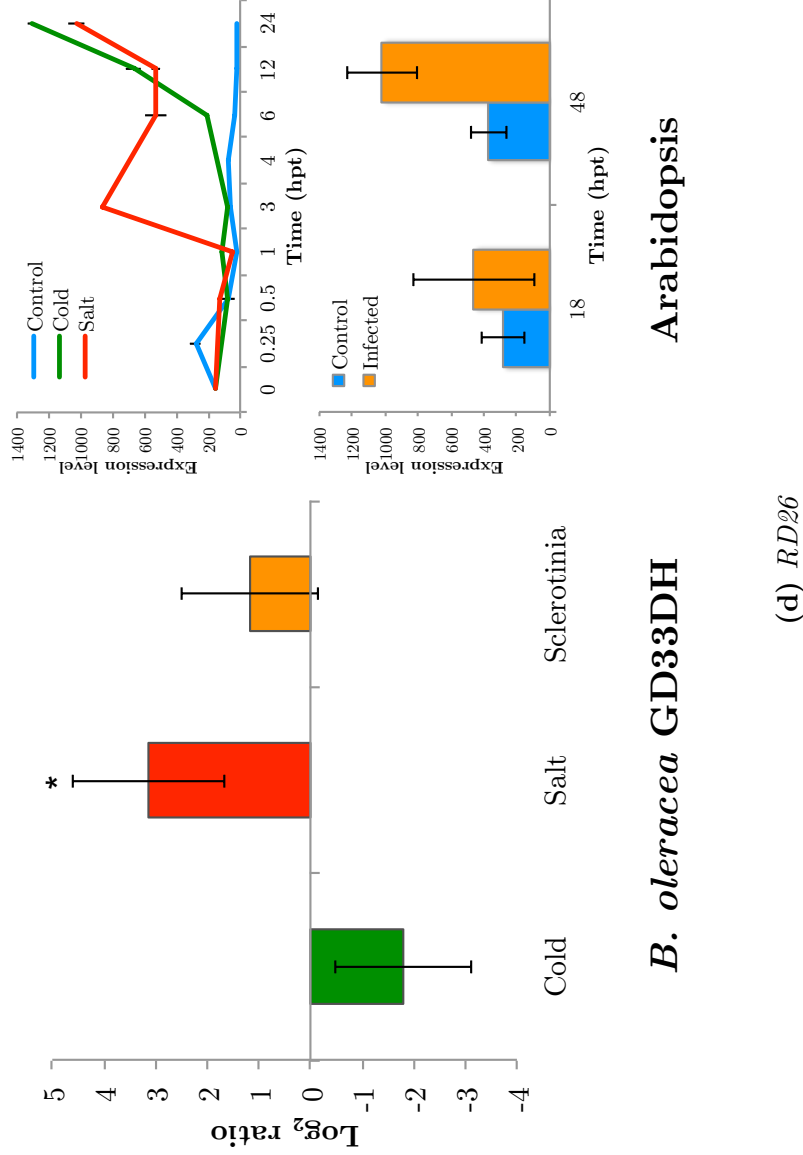


Figure 3.2: Experimental validation of the RD26 marker gene following stress treatment in *B. oleracea* GD33DH

Expression levels of the key marker genes (a) *PR1*, (b) *JAZ1*, (c) *COR15A* and (d) *RD26*, determined using qPCR on RNA isolated from stressed *B. oleracea* GD33DH leaf material (left) and in *Arabidopsis* using eFP expression data (right, Kilian et al., 2007). Experiments in *B. oleracea* GD33DH were carried out following cold stress (24 hpt, green), salt shock (1 hpt, red) and infection with *S. sclerotiorum* (48 hpt, orange). The error bars show the least significant difference (LSD) at 0.05, $n=3$. The log₂ ratio indicated on the y -axis represents the difference in expression between treatment and control. Significantly differentially expressed genes are marked with an asterisk (*). The *Arabidopsis* data is present as a time series for control (blue), cold (green), salt (red) and infection with *B. cinera* (orange). The x -axis indicates hpt. The y -axis represents the expression levels. The error bars indicate standard deviation. hpt, hours post treatment.

Levels of *PR1* (Pathogenesis-Related protein 1) were significantly elevated in response to infection with *S. sclerotiorum* in GD33DH (Fig. 3.2a, left), showing an induction of the basal level defence response. Salicylic acid, a hormone produced when the presence of a pathogen is detected in plant cells, induces the expression of *PR1*, which switches on the PR suite of genes. This basal level response against a broad array of pathogens, can be used to ‘prime’ the plant against further attack, known as ‘Systemic Acquired Resistance’, of which expression of *PR1* is often used as a marker (Laird et al., 2004). Enhanced expression of *PR1* was not seen following the abiotic stresses suggesting that priming does not occur under abiotic stress in GD33DH by infection with *S. sclerotiorum*. Data from Arabidopsis (Fig. 3.2a, right) indicate that *PR1* is not up-regulated in the leaves for the abiotic stresses but expression increases over time in response to *Botrytis cinera* infection (Winter et al., 2007), showing consistency with the response seen here in GD33DH.

Expression of *JAZ1* increased significantly following infection with *S. sclerotiorum* (Fig. 3.2b, left). *JAZ1* acts as a repressor of JA responsive genes by binding to the promoters of downstream genes and preventing transcription. Upon detection of JA, *JAZ1* is removed from the DNA and expression of JA responsive genes occurs (reviewed in Wasternack and Hause, 2013). In combination with the above result, this suggests that in response to *S. sclerotiorum*, the JA pathway is repressed in favour of the SA pathway in GD33DH. Although not significantly changed under cold conditions, *JAZ1* is down-regulated suggesting low-level activation of the JA signalling pathway under cold stress. This is contrary to the effect seen in Arabidopsis under cold stress (3.2b, right), as *JAZ1* accumulates to a high level compared to control suggesting differential regulation of stress response between species. It is likely that cross talk between hormone pathways accounts for the differential response between species under cold stress.

Expression of *COR15a* was significantly enhanced in response to cold stress in GD33DH and Arabidopsis (Fig. 3.2c). *COR15a* expression is activated by the CBF/DREBs (DREB1a, DREB1b and DREB1c) AP2/ERFs transcription factors (Wilkinson and Davies, 2002). The CBF proteins bind to the cis-elements in gene promoters to activate the expression of the CBF regulon, which includes the COR (cold responsive) genes. The COR genes have been implicated in freezing tolerance by the activation of multiple protective mechanisms (Gilmour et al., 2004; Wang and Hua, 2009). In GD33DH it can be seen that *COR15a* is significantly differentially expressed only in cold stress, not in the salt stress or infection with *S. sclerotiorum* (Fig. 3.2c, left). In Arabidopsis, *COR15a* was significantly up-regulated under both cold and salt stress (3.2c, right), however differences in experimental conditions and differential cross-talk between hormone signalling pathways may account for the difference between species.

Levels of *RD26* were significantly increased in response to salt shock but not to cold stress or biotic stress, showing how different signalling pathways can be used to fine tune

Treatment	Source	Transcripts
Cold, salt and <i>S. sclerotiorum</i>	GD33DH	112,130
Leaf and root	Cg-DFFS	218,503

Table 3.1: Metrics for genome guided assembly of RNAseq reads

Number of transcripts obtained from GD33DH under various stress conditions and Cg-DFFS leaf and root transcriptome data.

the response to stress (Fig. 3.2d, left). When *RD26* expression was analysed in Arabidopsis (3.2d, right) expression was up-regulated as a result of salt and cold stress in the leaves (Winter et al., 2007), again differences in experimental conditions and differential cross-talk between hormone signalling pathways may account for the difference between species.

In summary, the methods for stress treatments that were developed for GD33DH indicated a clear stress response in treated plants both visually and at the gene expression level. Quantitative PCR with key marker genes showed that the response to stress in GD33DH was appropriate with the associated marker genes being significantly differentially expressed in each of the stress conditions. The fact that different stress genes were differentially expressed in the treatments shows that G33DH is activating the appropriate hormone signalling pathways, as expected from studies carried out in Arabidopsis. Expression of these key marker genes gives confidence that the material collected was showing a response to stress and was therefore deemed suitable for RNAseq analysis.

3.3 Preparing a stress-specific transcriptome of *B. oleracea*

Despite the recent release of the TO1000 *B. oleracea* genome and transcriptome (Parkin et al., 2014), relatively little is known about the transcriptional changes that occur in this species during exposure to different stress stimuli. To address this question, an RNAseq experiment was designed to gather mRNA transcript sequence data from leaves of GD33DH plants exposed to both biotic and abiotic stress conditions. Not only did this reveal stress related transcripts and patterns of alternative splicing, it also provided additional GD33DH-specific sequence information. This was combined with leaf and root sequencing data from the CgDFFS (previously collected as part of the VeGIN project) to produce a new transcriptome database.

Metric	Raw Assembly	Cleaned Assembly
Total Transcripts	247,801	235,199
Mean Transcript Length	1,077	1,109
N25 (bp)	2,614	2,643
N50 (bp)	1,657	1,683
N75 (bp)	958	988
Total Length (bp)	266,881,587	260,793,546

Table 3.2: Assembly metrics for *de-novo* assembly of RNAseq reads

Raw assembly used transcripts greater than 200 bp, straight from the assembly software. The cleaned assembly used the metrics of the transcripts with contaminants removed.

3.3.1 RNAseq analysis of *B. oleracea* GD33DH under abiotic and biotic stress conditions

Following qPCR analysis of key stress genes (see Section 3.2), cold stress, salt shock and *S. sclerotiorum* infected samples were prepared for RNA sequencing, with three biological replicates per condition. Sequencing was carried out by the Genome Centre at Barts and The London School of Medicine and Dentistry in October 2013 using an Illumina HiSeq 2000. This resulted in the generation of 18 samples of 100 bp paired-end reads. Quality of the reads was assessed (Appendix C), whereupon it was discovered that there was a machine error on the R2 reads at the 18th base pair resulting in a miscalled base (N) at this location. In addition, some over-represented sequences mapping to known adaptor sequences were found within the reads. Therefore it was decided to preprocessing of the data through trimming and removal of over-represented adaptor sequences, as discussed in Methods (Chapter 2).

The paired end reads were aligned to the *B. oleracea* TO1000 genome (Parkin et al., 2014) using TopHat2 (Trapnell et al., 2012b) with an average 76.5% read alignment. Following alignment with TopHat2, transcripts were assembled using Cufflinks (Trapnell et al., 2010), generating detailed information on transcript variants compared to the TO1000 reference genome for each of the stress conditions. The transcript assemblies for each sample was merged to generate a comprehensive set of 112,130 transcripts expressed in GD33DH under various stress conditions (Table 3.1).

As part of a Defra funded Vegetable Genetic Improvement Network (VeGIN) project, leaf and root RNAseq data was generated for leaf and root material for members of the Cg-DFFS collection. This data was aligned to the TO1000 genome (carried out by Dr Jay Moore and Dr Yi-Fang Wang). As above, transcripts were resolved, generating a comprehensive set of 191,673 transcripts that were expressed in the leaf and root of various members of the Cg-DFFS collection (Table 3.1).

In addition, to capture sequence information that does not align to the TO1000

genome, reads were also assembled *de-novo* to identify additional transcripts, using the Trinity software (Haas et al., 2013). *De-novo* assembly is the joining of overlapping reads without the use of a reference genome to guide the assembly, avoiding potential bias. This produced an initial raw assembly of 247,801 contigs with a median length (N50 value) of 1,657 bp and average length of 1,077 bp (Table 3.2). This is shorter than the average length of an Arabidopsis gene (2,196 bp) (Wortman et al., 2003) but similar to the average transcript length of the TO1000 reference (1,042 bp). Arabidopsis genes have a longer average gene length due to the fact that there is more highly annotated genomic sequence for this model organism, whilst the genome of TO1000 was only recently sequenced.

To determine if transcripts from non-plant sources were contaminating the newly assembled transcriptome, a multi-genome blast database consisting of reference genomes from *Escherichia coli* (str. K-12 substr. MG1655), *Saccharomyces cerevisiae* (S288C) and *S. sclerotiorum* as well as plant reference genomes from *B. oleracea* (TO1000, v1), *B. rapa* (197) and Arabidopsis (TAIR10) was created and transcripts from the raw RNAseq assembly were searched against it. The majority of the contaminating transcripts originated from *S. sclerotiorum* due to the presence of fungal material in the samples from the infection process. A small number of contaminating sequences were found to be most similar to *E. coli* (80 sequences) and *S. cerevisiae* (9 sequences). The contaminating sequences were removed from the raw assembly to give a ‘cleaned’ assembly of 235,199 transcripts, with a higher N50 value (Table 3.2 and Fig. D.1 in Appendix D).

3.3.2 Generation of a non-redundant transcriptome database

All available *Brassica* transcript information was assembled together to generate a non-redundant transcriptome database. This consisted of three principle sets: the GD33DH genome guided transcript assembly, the Cg-DFFS genome guided transcript assembly (Table 3.1) and a cleaned *de-novo* assembly of the GD33DH stress RNAseq data (Table 3.2). A comprehensive transcriptome of 426,872 transcripts was produced. Redundant transcripts were removed by collapsing the transcripts into clusters at a 95% sequence similarity level and selecting a representative transcript for each cluster (Li and Godzik, 2006). This resulted in a new transcriptome of 213,110 non-redundant transcripts, which are available in additional datafile1 (Appendix I).

The next stage was to develop a microarray resource which could be used to capture subtle changes in gene expression in subsequent experiments.

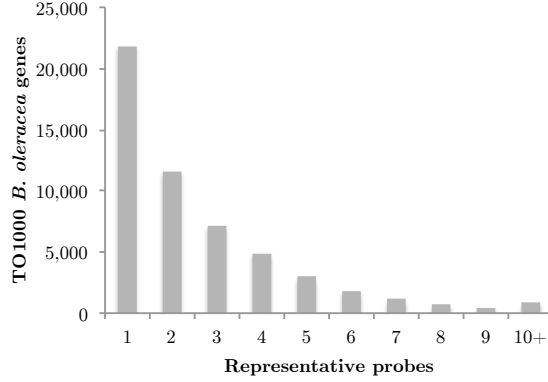


Figure 3.3: Number of probes per *B. oleracea* TO1000 gene ID present on the C-genome *Brassica* microarray

3.4 Microarray probe design

The *B. oleracea* genome has been shown to contain many highly repetitive regions, in addition to a highly duplicated genome resulting from recent WGD events (Parkin et al., 2014). In order to be able to distinguish paralogous sequences, it is important that the oligonucleotide probes are long enough to be able to resolve such closely related transcripts, therefore it was decided to use the Agilent Technology microarray platform, which uses 60-mer probes. The 4x180k microarray format consists of one glass slide containing 4 arrays of 180,880 features. For each array, the standard control grid requires 4,854 features, leaving 176,026 features available for probe design. The 213,110 cleaned, assembled transcripts were submitted to Agilent Technologies eArray web portal for gene expression probe design. The probes were 60mers, designed to be unique to each transcript in the transcriptome. Initially, the transcriptome output resulted in the design of 213,032 probes, of which 108,454 had cross-hybridization potential. Upon closer inspection of the probes, it was found that many were identical in sequence and so were removed. Following this, the number of probes in the design was reduced to 160,324, which was within the feature allowance. To use the full capacity of the array, 15,701 of these probes were randomly selected and duplicated within the design.

The transcripts were queried against the TO1000 gene models using BLASTn (with an E-value of $1e^{-05}$) to give the closest TO1000 gene for each transcript sequence. On the array, 127,553 probes were assigned a TO1000 gene, which resulted in 53,387 (90.1%) of the TO1000 genes appearing at least once. This resulted in some *B. oleracea* genes being represented by multiple probes on the microarray (Fig. 3.3). These multiple probes map to different parts of a gene and there would be useful in the analysis of differential splicing. In down-stream analysis it must be taken into consideration that some genes are represented by multiple probes, and these must be removed prior to further analysis. This can either be achieved by averaging out expression values or by taking the most highly

Species	Sequences
<i>A. thaliana</i>	229,085
<i>B. rapa</i>	135,037
<i>B. napus</i>	113,240
<i>B. oleracea</i>	2,564
<i>B. carinata</i>	211
<i>Camelina sativa</i>	106,402
<i>Capsella rubella</i>	54,394

Table 3.3: Species included in the pan-Brassicaceae BLAST database

Includes the number of sequences associated with each species used in the BLAST database. Sequences were downloaded from GenBank (nr database) and concatenated before a BLAST database was constructed.

expressed probe as a representative.

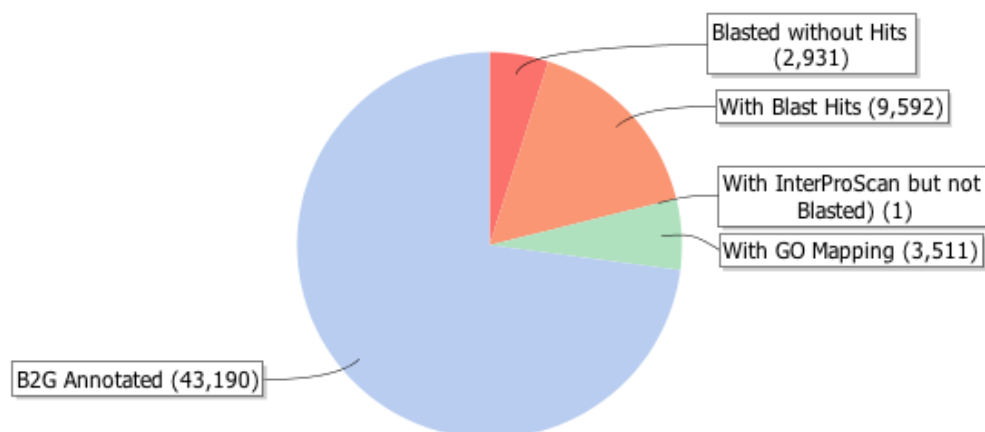
The TO1000 annotation provides information on Arabidopsis-*B. oleracea* orthology and the originating sub-genome of each TO1000 gene, from the more dominant LF sub-genome, to the Medium Fractionated (MF1) and Most Fractionated (MF2) sub-genomes (see Chapter 1). Based on this orthology, the largest proportion of probes mapped to the LF sub-genome (33,843 probes), followed by the MF1 sub-genome (25,639 probes) and finally the MF2 sub-genome (19,954 probes).

The remaining 32,770 probes were not annotated with a TO1000 gene, leaving 6,008 TO1000 genes unrepresented on the array. This included a large number of transposons, retrotransposons, hypothetical proteins, ribosomal proteins, TIR-NBS-LRR disease resistance genes and leucine rich repeat proteins.

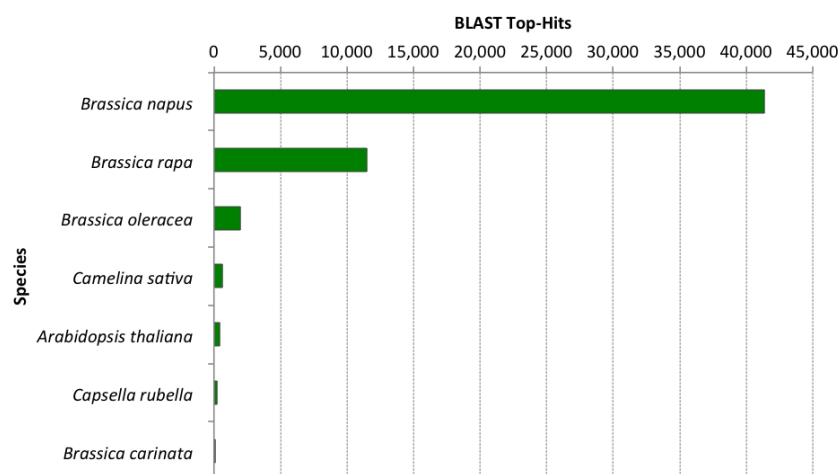
3.5 Annotating the *B. oleracea* TO1000 genome with GO terms

GO (Gene Ontology) is a structured language used to represent the properties of gene products using well defined terms and inter-related relationships, known as ‘GO terms’. Categories of GO terms include Biological Function, Cellular Component and Molecular Function. Each GO term has a precisely defined name, relationship and an associated ID. GO terms are structured as Directed Acyclic Graphs (DAGs), that is terms have a ‘root’ at the top and the nodes are connected by a relationship (e.g. ‘is a’ and ‘part of’), as one moves down the graph, the GO terms increase in specificity. Terms can have multiple parents, however they cannot link cyclicly and different paths can be taken down the graph (Ashburner and Lewis, 2002).

GO terms were assigned using the BLAST2GO software (Conesa and Götze, 2008), a GO term annotation software which infers GO terms using sequence similarity and an



(a) Number of annotated TO1000 genes



(b) Species top hit of TO1000 genes

Figure 3.4: BLAST hits of TO1000 gene sequences

The 59,225 TO1000 gene sequences were searched against a BLAST database of Brassicaceae sequences. (a) Number of annotated TO1000 genes and (b) the top species hit.

InterPro scan to identify suitable GO terms for genes without an annotation. The TO1000 genome was annotated with GO terms, firstly by creating a BLAST database containing all of the *nr* nucleotide sequences from various Brassicaceae species (see Table 3.3) with the aim of encapsulating as much sequence variation from the Brassicaceae family as possible. The TO1000 gene sequences were searched against this database (Fig. 3.4a) and the top hit for each gene was selected (Fig. 3.4b).

Using the species top hit and the protein domain information, GO terms were mapped to the genes using the TAIR and UniProtKB databases. Of the 59,225 gene models in the TO1000 genome, 43,190 genes were actually annotated with GO terms, the rest were not (Fig. 3.4a). The GO annotation for use with BiNGO can be found in additional datafile2 (Appendix I). A subset of these genes received no GO term annotation because they did not have a BLAST hit to any of the species in the BLAST database (2,931 genes), possibly because these genes are specific to TO1000 and are not found in other Brassicaceae species.

Searching for orthologs using a pan-Brassicaceae genome database, rather than a direct comparison back to Arabidopsis provides additional interesting information on the closest gene homology across the Brassicaceae. This gives further insight into the evolution of *B. oleracea* since its divergence from Arabidopsis. It also increases the coverage of the annotation by catching sequences that may not hit to Arabidopsis, due to rearrangements or sequence differences. The species showing the BLAST top hit was *B. napus* (Fig. 3.4b). This allotetraploid species is derived from the pairing of *B. oleracea* (C-genome) and *B. rapa* (A-genome) and has an AC genome, containing gene copies from both species (Chalhoub et al., 2014; U, 1935). That *B. napus* was the species with the most top hits is unsurprising, given that it contains the C-genome and was highly represented in the BLAST database (Table 3.3). The species with the next highest hit was *B. rapa*, again this species contains a large number of sequences in the BLAST database (see Table 3.3). Although it is the A-genome, *B. rapa* shares a high level of co-linearity between *B. oleracea*, with a small number of chromosomal rearrangements separating the two species (Parkin et al., 2014). The number of sequences matching *B. oleracea* was very low, however *B. oleracea* was very under represented in the *nr* database, since neither TO1000 genome sequence nor the *B. oleracea* var. *capitata* line 02-12 sequence (Liu et al., 2014) were present in the *nr* database at the time of analysis. The frequency of obtaining a top hit to an Arabidopsis sequence was very low, because there were species much closer to *B. oleracea* present in the BLAST database (see Table 3.3), indicating that there has been some sequence drift since the split between Arabidopsis and the *Brassica* sp. Despite being well represented in the BLAST database, there were few hits to *C. sativa* and *C. rubella*. These species are evolutionarily further away from *B. oleracea* than other Brassicaceae present in the BLAST database and could be considered out groups in the analysis.

GO terms are available at different levels of detail, for instance GO slim is a version of

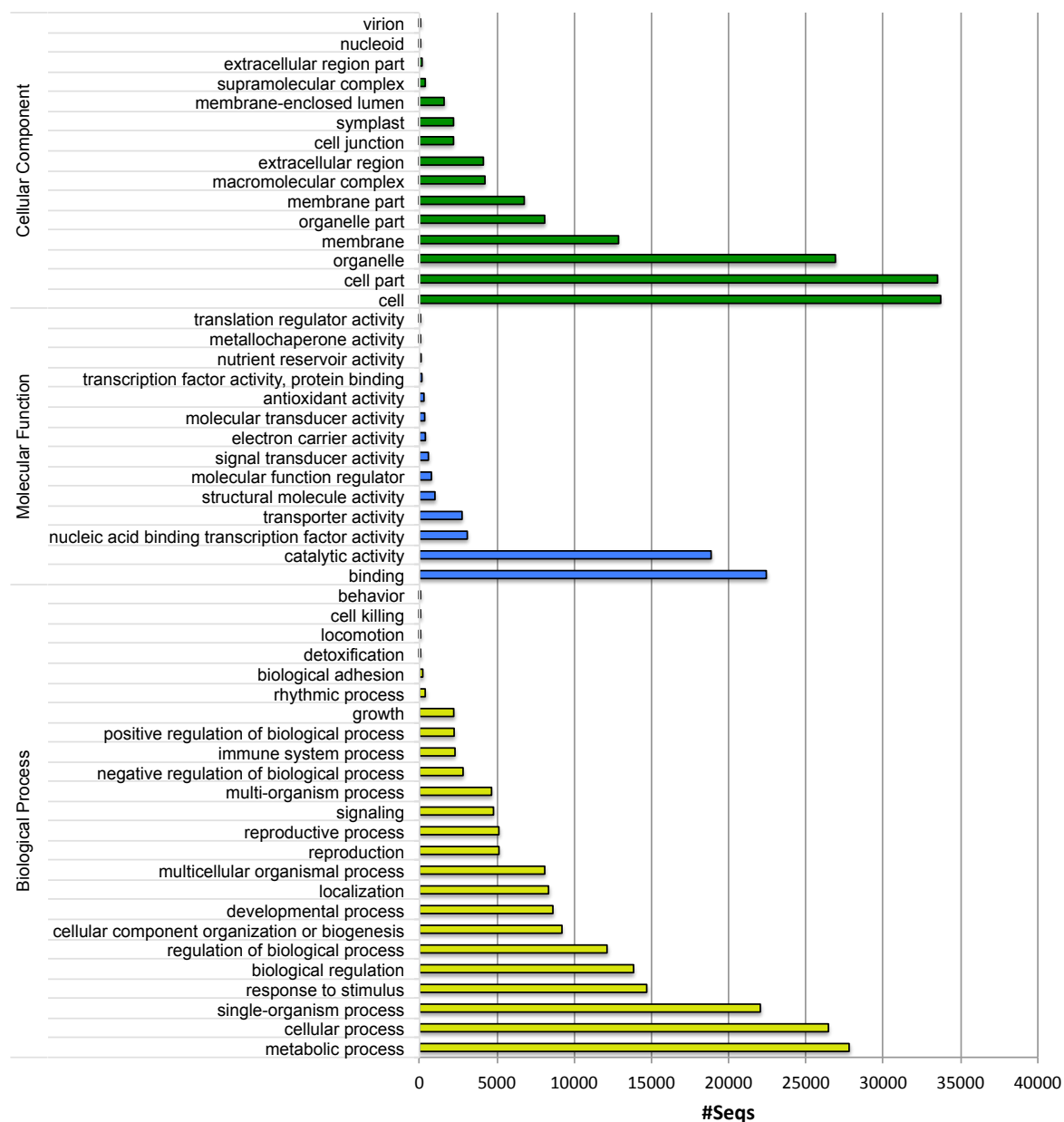


Figure 3.5: GO annotation of the TO1000 genome

Level 2 GO terms for Biological Process (green), Molecular Function (blue) and Cellular Component (yellow). *y*-axis gives the GO term for each category. *x*-axis shows the number of sequences annotated with the respective GO term.

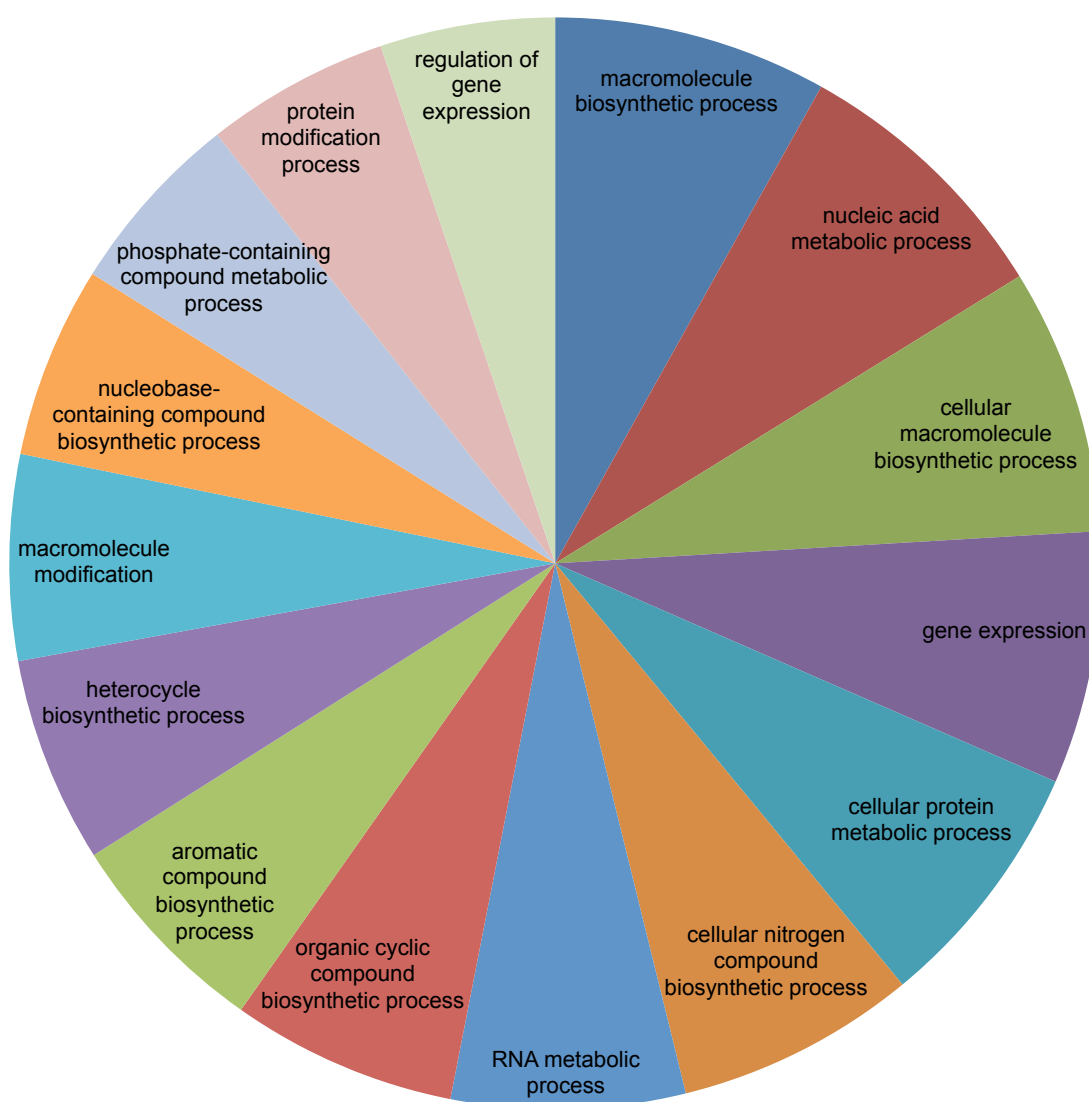


Figure 3.6: Top 20 level 5 GO annotations of the Biological Process category.

the ontology where more specific terms have been collapsed up into the more general parent terms. GO slims for all three categories including Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) were retrieved for *B. oleracea* TO1000 (Fig. 3.5). GO terms at this level could be considered useful when gaining a general picture of gene function. Alongside broad GO slim terms, more detailed GO terms, which provide a more specific annotation were also retrieved (Fig. 3.6). These GO terms may be important for certain types of analysis, for example when comparing highly specific stress related gene function.

3.6 Discussion

The chosen technology for the microarray was Agilent, which uses oligonucleotide synthesis for detection of mRNA. Agilent microarrays are renowned for producing high quality data, that is highly reproducible, reducing the need for large numbers of technical replicates (LeProust, 2015; Patterson et al., 2006). The 60-mer probes allow highly similar sequences to be resolved (Fenart et al., 2013), given the high level of duplication present in the *B. oleracea* genome this is advantageous in generating as much useful data as possible from the experiment.

A broad selection of source material, including *B. oleracea* GD33DH under abiotic and biotic stress conditions and leaf and root transcriptome sequence of C-genome *Brassica* species permits the discovery of transcripts which have not been previously recorded in *B. oleracea*, as indicated by the large number of transcripts without orthology to the TO1000 genome.

A transcriptome was assembled containing 213,110 non-redundant C-genome *Brassica* transcripts. Subsequent 60-mer probe design resulted in the design of 160,324 probes representing, 90.1% of the *B. oleracea* genome. The remaining 9.9% transcripts were too repetitive to be resolved at the probe level and included a selection of transposons, which are over-represented in *B. oleracea* (Town et al., 2006) and also highly repetitive gene families such as the TIR-NBS-LRR disease resistance genes (Meyers et al., 2003).

Transcriptional profiling using microarrays is an extremely powerful technology for identifying genes involved environmental response and adaptation. CATMA microarrays have been essential in the determining Arabidopsis response to stress conditions, and the identification of key regulatory genes using gene network inference (Bechtold et al., 2016; Breeze et al., 2011; Windram et al., 2012). The C-genome *Brassica* microarray will be useful in determining *Brassica* genes that are important in the response to stress conditions and in the leaf and root, thus having a multitude of potential uses beyond this thesis.

In addition, the GO annotation of the *B. oleracea* genome is a valuable resource, without which GO analysis would take place using the closest Arabidopsis ortholog.

Having a direct *B. oleracea* annotation will be useful in interpreting the output of the microarray and also in other experiments involving the use of the TO1000 genome that have been carried out in this thesis. When using GO annotation, it must be considered that annotations are inferred from homology to genes and protein domains with previously annotated functions and existing annotations are not likely to be complete (King et al., 2003; Pinoli et al., 2015). For most sequences organisms, only a small selection of known genes have been functionally annotated and those that are annotated are most likely to be related to developmental process and environmental responses due to the nature of scientific research.

3.7 Chapter Summary

A new microarray has been designed for *B. oleracea* using new genomic and transcriptomic data that has recently been made available. The transcripts present on the microarray were assembled from the published *B. oleracea* TO1000 genome, RNAseq data from stress experiments carried out in *B. oleracea* GD33DH and also leaf and root transcriptome data from the C-genome DFFS collection of wild *Brassica* species. Sequence similarity was used to assign the closest TO1000 genome model to each of the probe transcripts present on the array. In addition, GO annotation of the *B. oleracea* TO1000 genome was carried out, useful for functional analysis of groups of genes. The outcome is a new transcriptome and annotation for C-genome *Brassica*, containing previously unidentified transcripts as well as associated GO terms for the transcriptome. The microarray and GO annotations will be used later on in this thesis to measure and analyse gene expression for a high-resolution time-series experiment carried out in GD33DH.

Chapter 4

High-resolution time series transcriptomics of salt shock in *Brassica oleracea*

4.1 Chapter Overview

Time-series transcriptomic experiments, in which sampling of plant material occurs at close, regular intervals across a specified time period in a highly replicated manner, are very valuable in capturing subtle changes and fluctuations in gene expression, producing meaningful biological knowledge. This method has been carried out extensively in *Arabidopsis* to decipher changes in gene expression in response to pathogenic species such as *Botrytis cinera* (Windram et al., 2012) and *Pseudomonas syringae* (Lewis et al., 2015), abiotic stress conditions such as drought stress (Bechtold et al., 2016) and during developmental processes such as senescence (Breeze et al., 2011). Though analysis of crop plants to high-salt conditions has been carried out in various species such as *B. napus* (Liu et al., 2015a), cotton (Xu et al., 2013a) and rice (Walia et al., 2006b), analysing gene expression changes as a high-resolution time-course is rarely carried out in crop species, due to the high expense of the experiment and lack of extensive genomic resources available for crop species.

In this chapter, the aim was to investigate the complex physiological and genetic mechanisms involved in the early stages of salt shock (0-36 h) in *B. oleracea* GD33DH. In order to achieve this aim, a high-resolution time series analysis was carried out in which gene expression was measured at 2 h intervals for a total of 36 hours using microarrays (as designed in Chapter 3).

Since plants are unable to move away from high stress environments, large scale changes in gene expression are an essential part of the protective mechanisms that plants use to tolerate stressful conditions. This high-resolution transcriptomic time series experiment

revealed 7,141 genes which were differentially expressed in response to salt shock. Analysis of leaf Na^+ concentration, ABA biosynthesis genes, diurnal gene expression, TF families, ion transporters and GO terms of differentially expressed genes have been used to piece together a chronology of the early salt shock response.

The results described in this chapter greatly enrich the existing information on salt response mechanisms of *B. oleracea* and provide numerous candidate genes for further analyses and for potential manipulation to improve the salt tolerance of *Brassica* crops.

4.2 Results

A high-resolution time series experiment was performed sampling salt treated and control *B. oleracea* GD33DH leaves every two hours over 36 hours. Global gene expression of the time series was profiled to identify transcriptomic changes that occur in GD33DH under salt shock. A high concentration saline solution (250mM; control plants were watered with deionised water. See Chapter 3 for preliminary experiments) was applied to GD33DH grown on compost to ensure transpiration was not limited. Three independent biological replicates of leaf #5 sampled from separate GD33DH plants were harvested at each time point for each treatment.

4.2.1 Physiological effects of salt shock in *B. oleracea* GD33DH

Within 2 hpt (hours post treatment), visible wilting of the leaves was seen as plants lost their turgidity, which was regained by 6 hpt. This response is typical of an osmotic stress, in which the excess Na^+ ions at the roots lower the water potential within the root cells and water is drawn out of the plant (Downton and Millhouse, 1983; Kumar et al., 2009). One of the mechanisms plants employ to protect themselves from the effects of high Na^+ is to maintain a high concentration of K^+ relative to Na^+ ($\text{K}^+:\text{Na}^+$ ratio), preventing ionic stress caused by an excess of Na^+ in the cytoplasm. In order to determine whether excess Na^+ ions are entering the root and being transported to the shoot, or whether GD33DH is successfully able to exclude excess Na^+ through maintenance of a high $\text{K}^+:\text{Na}^+$ ratio in the shoot, Inductively Coupled Plasma Mass Spectrometry (ICP-MS) of freeze dried GD33DH leaf material was carried out in order to quantify the Na^+ and K^+ content of the leaf. Samples were taken at 0, 6, 12, 24 and 36 hpt and the ICP mass spectrometry was carried out by Almustapha Lawal (University of Warwick), as described in Methods (Chapter 2).

From the results of the ICP mass spectrometry (Table 5.7) it can be seen that within 2 hpt, the $\text{K}^+:\text{Na}^+$ ratio decreases from 13.0 at 0 hpt and with minor fluctuations the lowest ratio of 5.4 is reached by 24 hpt. After 24 hpt, protective mechanisms become effective

Time point (hpt)	K ⁺ μ /g	Sig.	Na ⁺ μ /g	Sig.	K ⁺ :Na ⁺
0	1305.6 \pm 253.3	NA	102.2 \pm 9.5	NA	13.0
2	1015.1 \pm 685.0	NA	123.6 \pm 24.1	NA	8.5
6	1271.0 \pm 167.4	NA	123.4 \pm 8.8	*	10.3
12	1428.2 \pm 1005.8	NA	141.0 \pm 30.4	NA	9.7
24	764.7 \pm 214.7	NA	138.9 \pm 22.5	NA	5.4
36	1980.6 \pm 388.7	NA	145.6 \pm 7.4	**	13.7

Table 4.1: Mineral analysis of *B. oleracea* GD33DH leaves in early stages of salt shock

The graph shows the K⁺:Na⁺ ratio at selected time points. The table shows the mean K⁺ (\pm standard deviation) and mean Na⁺ (\pm standard deviation) mineral content and significance compared to time point zero (as determined using a *t*-test, NS=Not Significant, * $p<0.05$, ** $p<0.01$ and *** $p<0.001$) of *B. oleracea* GD33DH leaves undergoing salt shock at 0, 6, 12, 24 and 36 hpt. Mineral content was measured using ICP mass spectrometry (carried out by Almustapha Lawal, University of Warwick).

and the K⁺:Na⁺ ratio and is returned to time point zero levels of 13.7 by 36 hpt (Table 5.7). This drop in K⁺:Na⁺ ratio may be caused by increase in the Na⁺ concentration within the shoot which starts to rise at 2 hpt, is significantly different from zero levels by 6 hpt ($p<0.05$), and plateaus at 12 hpt (Table 5.7). This rise and subsequent plateau in Na⁺ concentration could be because the Na⁺ ions reach the shoot either through the bypass flow mechanism or through xylem loading before homeostasis is established at 12 hpt. This suggests that the initial osmotic stress seen by GD33DH following salt shock is under control by 12 hpt, and by 36 hpt the plant has managed to accumulated enough K⁺ ions to return to a similar K⁺:Na⁺ ratio at the zero time point. Further experimentation with a larger number of replicates and time points would be necessary to draw firm conclusions, as the data showed a large amount of variation.

4.2.2 Experimental design of the time-course microarray experiment

Hybridizing each treatment sample to its corresponding control (i.e. hybridizing treatment and control samples from time point 1 together) onto the same array would result in an unconnected design from which it would not be possible to make robust comparisons across time. In order to maximise the comparative power of the experiment, the RNA prepared from the experiment was hybridized in a randomized loop design. The loop design for the time series experiment was developed by Andrew Mead (Rothamsted Research, UK), and was designed to incorporate the following constraints:

- Two conditions - ‘Treatment’ and ‘Control’.
- Nineteen time points, including a time zero time point.

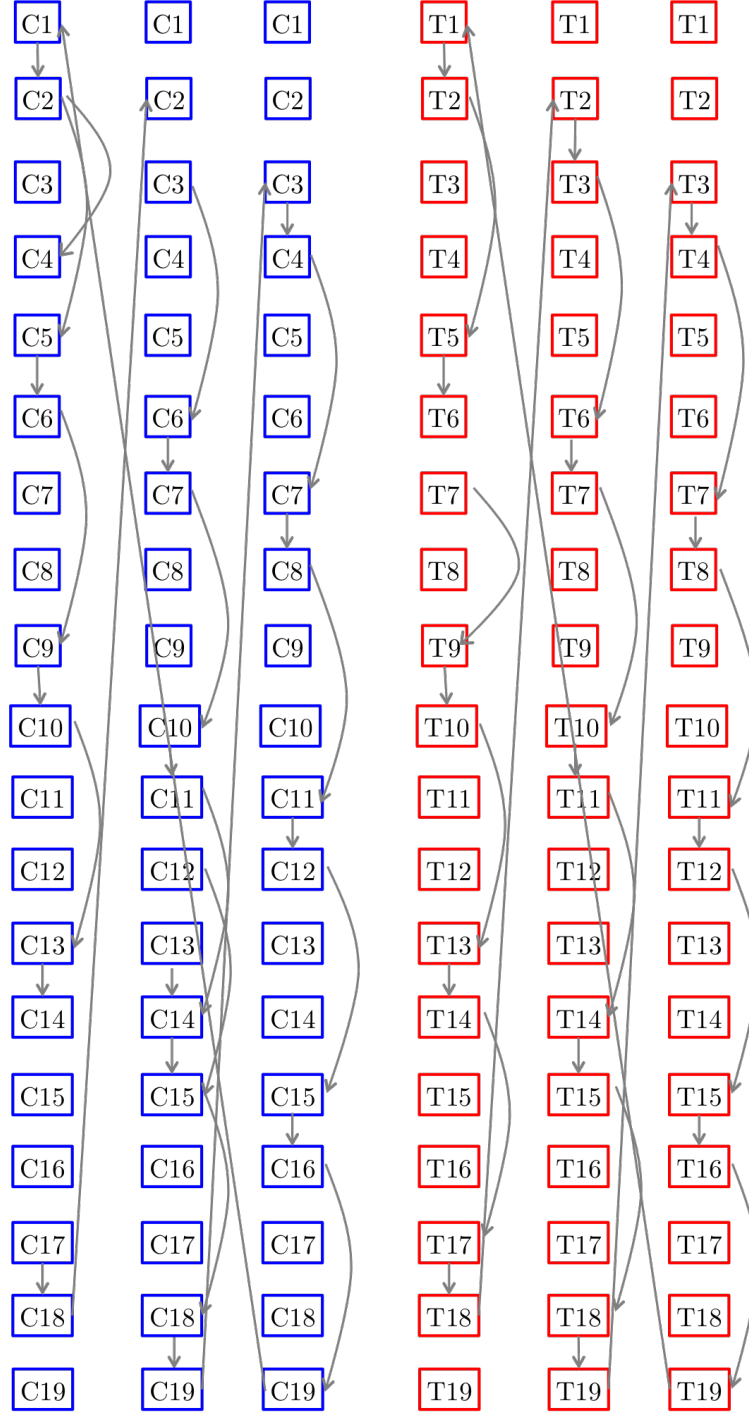


Figure 4.1: Loop design to link biological replicates over time

The first part of the loop design, showing all connections within biological replicates over time. Samples labelled in Cy5 lie at the head of the arrows and Cy3 at the tail. Each row is a different time point separated by 2 hpt intervals. Control ('C') samples are labelled in blue, and salt treated samples ('T') are labelled in red.

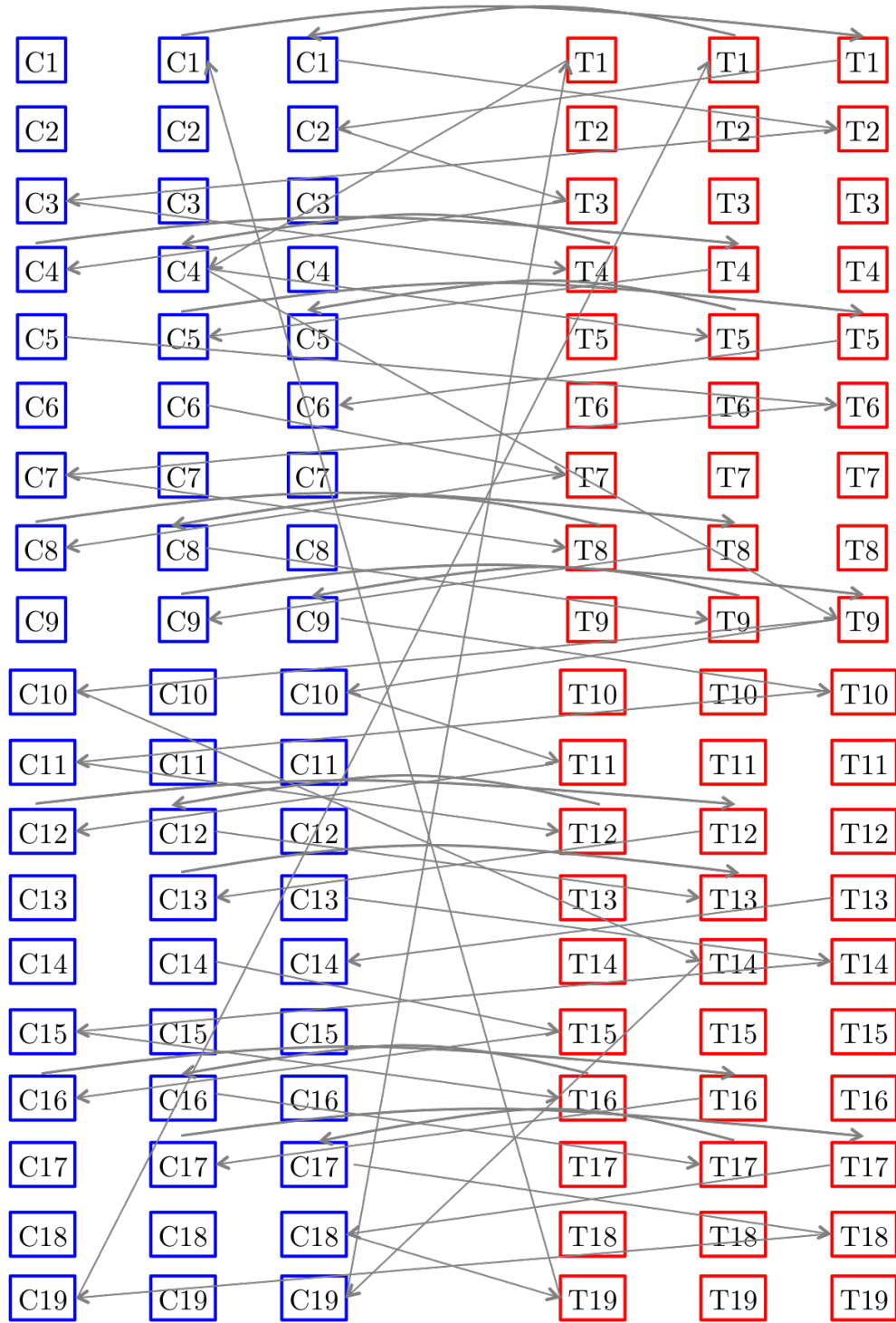


Figure 4.2: Loop design to link treatment and control conditions

The second part of the loop design, two natural loops showing all connections between conditions. Samples labelled in Cy5 lie at the head of the arrows and Cy3 at the tail. Each row is a different time point separated by 2 hpt intervals. Control ('C') samples are labelled in blue, and salt treated samples ('T') are labelled in red.

- Three biological replicates ('BioReps') per condition.
- Two technical replicates for each BioRep in the form of a dye-swap i.e. RNA from each sample is to be labeled with both Cy3 and Cy5 dye.

This resulted in an interwoven loop design as follows:

- Two time loops to link the ends of each biological replicate to the start of the next. This consisted of 29 arrays for each condition, totalling 58 arrays (All connections are shown in Figure 4.1).
- Two natural loops to link between conditions and between biological replicates. Some of these comparisons are within a time point and are between adjacent time points but all comparisons are between the two conditions. This consisted of 28 arrays for each loop, totalling 56 arrays (All connections are shown in Figure 4.2).
- Extra connections between loops to strengthen the design (6 arrays, not shown).

There were 120 microarrays in total, which were labelled and hybridised as described in the methods section (Chapter 2). The microarrays were scanned twice, with lasers of different wavelengths to excite the Cy3 and Cy5 fluorophores that were hybridized to the platform. The brightness of each individual spot was measured, and the analogue signal was translated to a fluorescence intensity value, which was used in all of the following downstream analysis.

4.3 Processing high-resolution time-series transcriptomic data

4.3.1 Data transformation and normalisation of *B. oleracea* GD33DH time series gene expression data using MAANOVA

The software package MAANOVA (MicroArray ANalysis Of VAriance; Churchill, 2004; McHattie, 2011; Wu et al., 2002) was used for the statistical analysis of the gene expression data generated from the microarrays. This program can handle time-series data and is used for assessing data quality, applying data transformations, fitting ANOVA models to estimate relative gene expression levels and carrying out F -tests for differential expression analysis.

Using background corrected data for microarray analysis increases variability of the log ratios at lower intensities (Churchill, 2004), thus, due to the high quality and consistency between each of the Agilent arrays, it was decided to use non-background corrected data

for subsequent analysis. Data was \log_2 transformed when loaded into MAANOVA to make it normally distributed. A \log_2 transformation redistributes the data across the entire intensity range rather than it being squeezed up at the lower end of the scale. Using the data on a log scale also makes comparisons between up-regulated and down-regulated transcripts simpler, as a single unit of change is the same in both directions.

4.3.2 Quality Assessment and LOWESS Transformation

Prior to model fitting, the quality of the arrays was checked pre- and post-normalisation using the RI plot ('Ratio x Intensity' plot) option in the MAANOVA package. A scatterplot of the \log_2 ratios of each dye for each individual probe plotted against the sum \log_2 intensity was generated for each array (120 in total) to check for technical and systemic bias and inconsistencies in hybridization. By minimising the systemic variation within the data, it is possible to identify true biological variation within the data. The underlying assumption of gene expression microarray analysis is that most genes will not change in expression between the two samples, thus their \log_2 ratio will lie along the zero mark on the y -axis. Transformations of the data rely on this assumption. A small number of genes will be differentially expressed, these will have more extreme \log_2 ratios, causing the 'scatter' effect that is seen in RI plots.

RI plots for each array were generated both pre- and post-normalisation (see Fig. 4.3). Due to a bias in dye incorporation during the labelling step and/or different responses of the dyes to the laser activation, a curve towards one of the channels was seen in the pre-normalisation plots (Fig. 4.3a). Therefore, a LOWESS normalisation was applied to the data to correct the dye bias (Fig. 4.3b and c). LOWESS is a curve fitting transformation which fits a regression line to the \log_2 ratio by a locally weighted least squares method, shifting the raw data so that the $\log_2(\text{Green})$ and $\log_2(\text{Red})$ are proportional to each other. The LOWESS normalisation was carried out in two stages, first a global LOWESS followed by a regional LOWESS. The global LOWESS is an intensity based adjustment, which aims to smooth the scatter plot of ratio verses intensity. After this transformation the regression line should pass through zero, straightening up the scatterplot (Fig. 4.3b). The regional LOWESS is a further transformation applied to the data to remove spatial biases on the array (Fig. 4.3c). This transformation takes into account the location of each probes on the array, by row and column providing spatial awareness to the regression. Due to the Agilent arrays being of extremely high quality compared to in house printed arrays, this second transformation had little effect on the data, as seen in Figure 4.3c. The clusters of probes (indicated by the red circle) seen outside the normal range of the scatter on all plots in Figure 4.3 are the positive and negative control set, which were included in the Agilent design. These positive and negative controls show predictable signal intensities and are used to check the quality of the array and the hybridization process. Negative

controls are also important for the background subtraction algorithms. Although these probes were included in Figure 4.3, they were removed prior to the LOWESS normalization so as not to affect the placement of the regression line.

4.3.3 Fitting a mixed model to gene expression data

A mixed ANOVA model was fitted to the data. This allowed the identification, isolation and removal of the sources of variation, resulting in estimated data for each treatment and time point.

In order to fit a model to the salt shock microarray data, the terms of the model must first be defined. Variation in data can be caused by a number of reasons. Technical variation caused by variation within the printed arrays, variation between the labelling and the hybridization process is not controlled by the user, thus defining these terms as ‘random effects’ is imperative to the model design. Sources of variation that can be controlled by the user, such as treatment, time point, number of biological replicates are known as ‘fixed effects’. By including both fixed effects and random effects in the model, sources of variation that are of interest, such as the effect of the treatment over time, can be separated from sources of random variation which are not of interest to the end analyses.

A mixed model of both fixed and random terms was applied to the data, incorporating the following random terms:

- *Dye* - Effect contributed by using different dyes (Cy3 and Cy5).
- *Array* - Effect contributed by using different arrays for hybridisation.

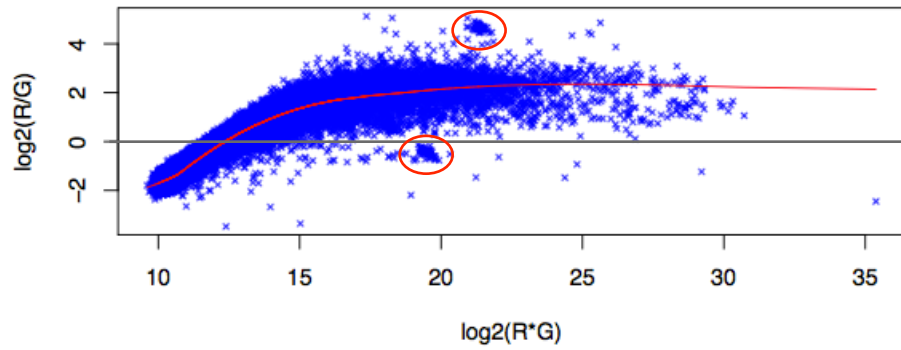
The fixed terms are defined in the experimental design:

- Treatment - Treatment received by the samples, either Control or Treated.
- Time - The time at which the samples were collected.
- Biological replicate (BioRep) - Three biological replicates were taken at each time point for each condition.

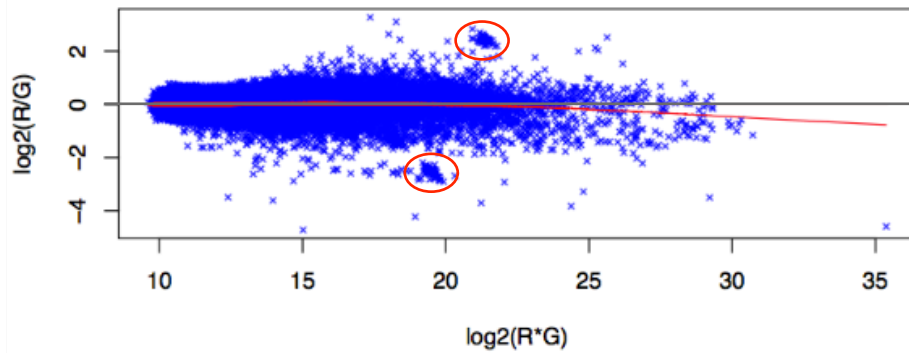
The model formula applied to the data was:

$$\sim \textit{Dye} + \textit{Array} + (\text{Treatment} * \text{Time})/\text{BioRep}$$

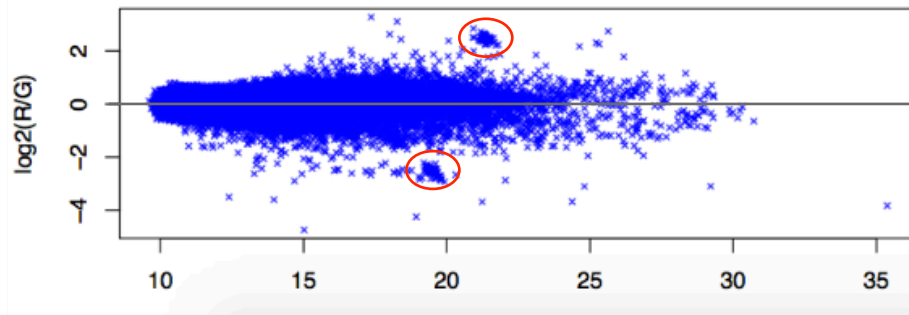
where terms written in *italics* are the random terms. The Time and Treatment terms were modelled separately but also with an interaction between the two terms, as there were both time and treatment dependent effects over the course of the experiment. By



(a) Before gLOWESS



(b) After gLOWESS, before rLOWESS



(c) After rLOWESS

Figure 4.3: RI plots before and after the application of global and regional LOWESS transformation

Microarray #6 under going LOWESS normalisation in two stages - global LOWESS (gLOWESS) (a) and regional LOWESS (rLOWESS) (b) to produce normalized data (c). The y -axis is the \log_2 ratio($\log_2(\text{Red}/\text{Green})$) and the x -axis is the \log_2 intensity ($\log_2(\text{Red} \times \text{Green}/2)$). The red line represents the LOWESS fitted curve and the grey line passes through $y=0$. The red circle represents control sets.

including this interaction, terms that were dependent on both the time and the treatment received were captured rather than lost to the measurement error. BioRep is nested within the Treatment * Time model, as each BioRep is only comparable with those at the same time point. The model was applied to the time series data to extract meaningful biological data.

4.3.4 Extracting predicted gene expression data from MAANOVA

Since each source of variation was captured within the mixed model, predicted expression profiles for each transcript could be extracted from the mixed model output with as many fitted terms included as required. Following extraction of the predicted data with the desired number of terms, further normalisation of the data takes place depending on downstream analysis:

- Raw expression values for each treatment at each time point on the original expression scale.
- Mean centred expression values - expression profile lies along mean expression of zero. This is useful for comparing between transcripts where relative expression is not important.
- Mean centred, standard deviation normalised values - all values are mean centred and the data are transformed so that the standard deviation is zero. This is useful for clustering based on shape rather than expression profile.

Data was extracted for both treatment and control for each time point individually, by including the BioRep term in the extraction or excluding the term to extract the data with the biological replicates combined. The standard error was also calculated for each time point, using the standard deviation of the variability across biological replicates.

4.4 Analysis of high-resolution time-series transcriptomic data

It is important to clarify the nomenclature used in the following analysis for the results of the time series microarray experiment. ‘Transcript’ is used to refer to the target mRNA sequence which has bound to its complementary probe. By querying the transcripts against a TO1000 BLAST database, the transcripts were mapped back to the TO1000 transcriptome to assign each one a Bo gene ID. Some transcripts did not map back to the TO1000 transcriptome and as such were not assigned a Bo gene ID. Where a Bo gene ID is used in downstream analysis, it must be kept in mind that this refers to a transcript

Test	F_s	FDR correction
Time	94,355	Unadjusted
Treatment	172,036	Unadjusted
Time x Treatment	367	Unadjusted
Time	23,370	StepDown
Treatment	94,233	StepDown
Time x Treatment	42	StepDown
Time	81,063	StepUp
Treatment	158,709	StepUp
Time x Treatment	367	StepUp
Time	94,355	Adaptive
Treatment	172,036	Adaptive
Time x Treatment	367	Adaptive

Table 4.2: The number of differentially expressed transcripts identified by F -tests, followed by different methods of FDR correction

Number of differentially expressed transcripts for the F -test terms at $p < 0.05$. Different methods of FDR correction were StepDown, StepUp and Adaptive (Benjamini and Hochberg, 1995, 2000; Benjamini and Liu, 1999)

which has mapped back to that Bo gene ID from the TO1000 transcriptome. The Bo gene IDs are useful in certain downstream analysis such as orthology to Arabidopsis, GO term analyses and to provide context to the transcripts.

4.4.1 Identifying differentially expressed transcripts

Differentially expressed transcripts are those that are expressed significantly differently in the treatment conditions compared to the control. Two methods were used to detect differentially expressed transcripts. Differentially expressed transcripts were found using two methods specifically adapted to time series microarray data - F -tests from the MAANOVA package (Churchill, 2004; Kerr and Churchill, 2001; Wu et al., 2002) and a locally adapted Gaussian Process Two Sample test (GP2S) (Stegle et al., 2010; Windram et al., 2012). The intersection of differentially expressed transcripts identified by the two methods was used to select the final list of differentially expressed transcripts, ensuring robust a select with minimal false positives.

F-tests

The F -tests were carried out on the fitted model for the fixed terms ‘Treatment’, ‘Time’ and ‘Treatment x Time’. The F -statistics were calculated by:

$$F = \text{explained (term) variance} / \text{unexplained (error) variance}$$

Transcripts are considered differentially expressed if their variance of the selected term is greater than the background noise. The variability between biological replicates can be included in the analysis to correct the F -statistic. This was done by carrying out F -tests on all of the fixed terms and recalculating new F -statistics with the denominator changed to the variance of the biological replicates. New p -values associated with this statistic are obtained from the F -distribution, a p -value <0.05 indicates that the transcript is differentially expressed.

A large proportion of the transcripts (106,057; 60.3%) were considered differentially expressed following F -test analysis (see Fig. 4.4). Due to the large number of tests being carried out the data, the p -values suffer from Type 1 error, the presence of false positives which have a p -value <0.05 . It is important to lower the p -value in order to control for false positives, whilst still capturing true positives. This adjustment is often seen as a trade-off, as the more strict the adjustment, the more false negatives will be present. Various methods of multiple testing correction were applied to the data to reduce the presence of false positives in the data (see Table 4.2). Methods included the StepDown (Benjamini and Liu, 1999), StepUp (Benjamini and Hochberg, 1995) and Adaptive (Benjamini and Hochberg, 2000) approaches.

Using the StepDown method of FDR correction (Table 4.2), an extremely large number of the transcripts were identified as differentially expressed for the ‘Treatment’ term (82,685; 46.9% of the total transcripts), the category of interest. These show expression profiles over 36 hpt that are significantly different in the salt-treated compared to the control samples, suggesting that a large alteration of gene expression takes place following salt shock treatment in GD33DH. In Figure 4.4, in the ‘Treatment’ category, the exemplar plot is the expression profile of a transcript mapping to Bo2g047740.1 which encodes a MATE efflux protein, a transporter involved in restoring the homeostatic balance after disruption with salt shock. The expression of this gene increases within 2 hpt, whilst the control expression profile remains relatively unchanged throughout the time-course, suggesting an early and constant role in the response to salt shock for this transcript.

For the ‘Time’ term there were 11,823 (16.7%) differentially expressed transcripts. Due to the oscillating nature of these expression profiles, it is likely that many of the transcripts are under regulation of the circadian clock. In Figure 4.4, the transcript used as an example expression profile for the ‘Time’ term maps to Bo5g002760.1, encoding *LATE ELONGATED HYPOCOTYL (LHY)*, a key component of the morning loop of the circadian clock (Pokhilko et al., 2012). The LHY component of the clock oscillates over a 24 hour period. When comparing for differences between treatment and control, expression of this transcript does not change except at the last time point (36 hpt) where expression of the treated sample increases compared to control.

Of the intersection of the two terms, 11,547 (6.5%) of the transcripts were considered

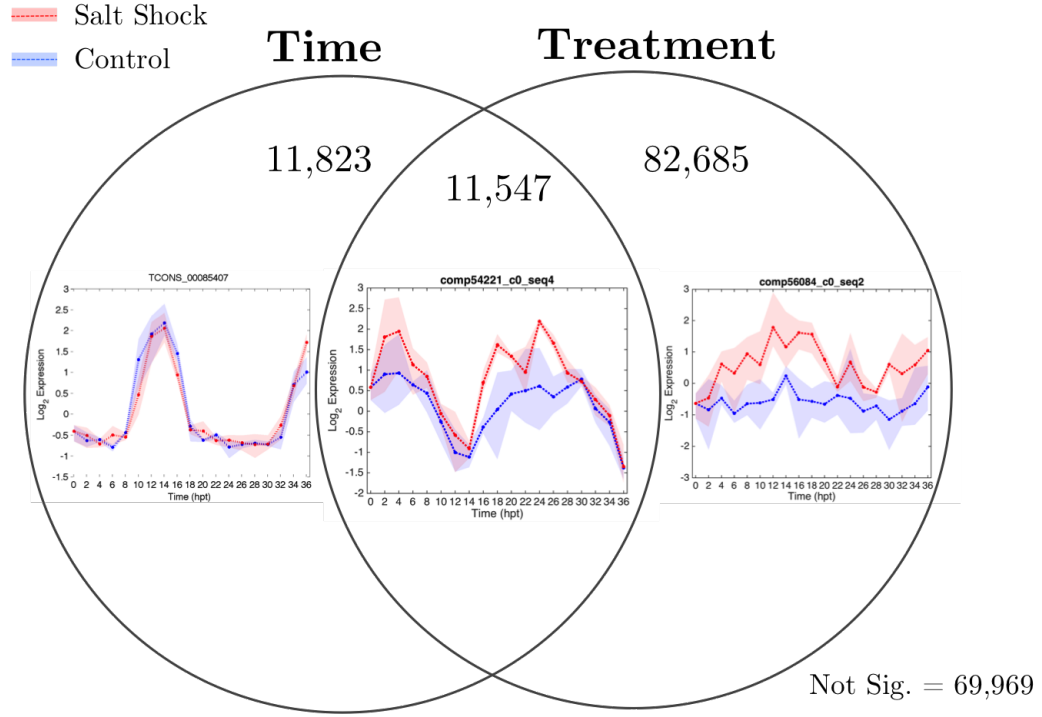


Figure 4.4: Venn diagram showing the number of transcripts with significant differential expression using F -Tests in MAANOVA

For each fixed term of the MAANOVA model, ‘Time’ and ‘Treatment’ were assessed for differential expression relative to the biological variation using an F -Test. The number of genes with positive test statistics (FDR corrected) for each combination of terms (‘Time’, ‘Treatment’ and ‘Treatment and Time’) is given along with exemplar plots. Red are salt shock treated and blue are control expression profiles, the shaded areas related to the minimum and maximum expression values. Time along the x -axis, Log₂ Expression on the y -axis. The transcript representing the ‘Time’ category maps to Bo5g002760.1, which encodes LHY; the ‘Time x Treatment’ category maps to Bo4g190900.1 (encoding ERD15); the ‘Treatment’ transcript maps to Bo2g047740.1 which encodes a MATE efflux protein family member.

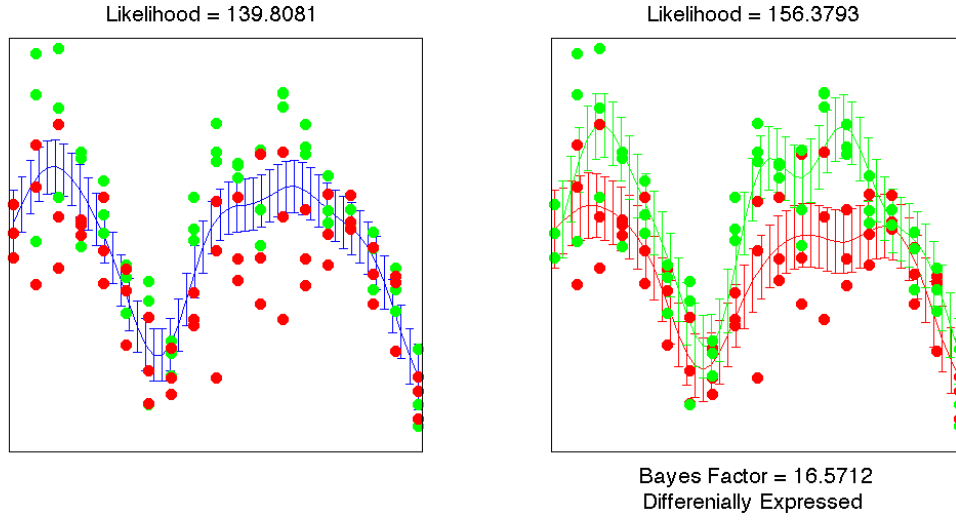
differentially expressed according to both treatment and time factors. Expression profiles may show diurnal expression and also difference caused by the effect of the treatment. Here, the exemplar plot used to represent the interaction between the ‘Time’ and ‘Treatment’ term is a transcript mapping to Bo4g190900.1, which encodes *EARLY RESPONSE TO DEHYDRATION15 (ERD15)*. *ERD15* is an ABA inducible gene, which negatively regulates the ABA dependent pathway in response to abiotic and biotic stresses (Kariola et al., 2006). Differential expression of this transcript at 2 hpt suggests an early role for this transcript, as is reported in the literature (Kiyosue et al., 1994; Wang et al., 2003).

Gaussian Process Two Sample

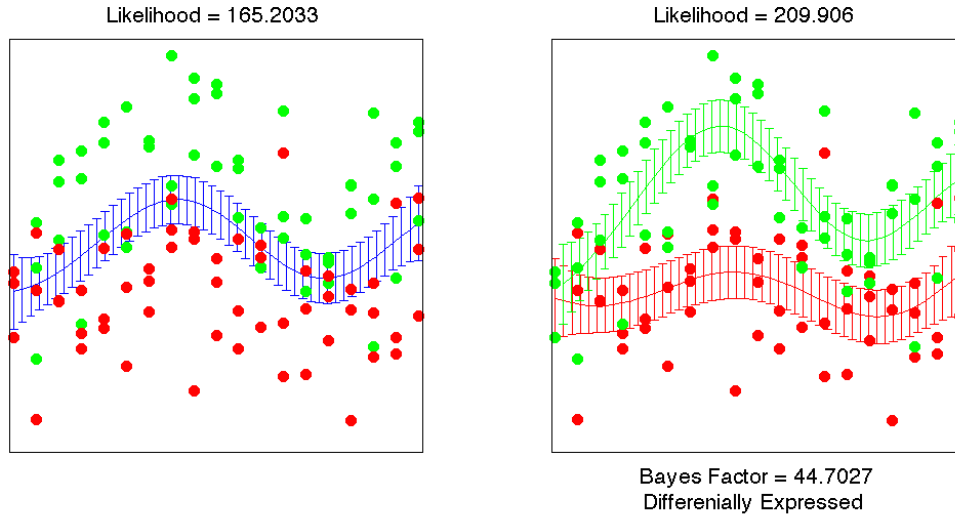
An additional method of detecting differentially expressed transcripts is the GP2S test, which allows the user to detect time-dependent differential expression, calculating the time point at which a transcript is first differentially expressed (Stegle et al., 2010). The rationale behind the GP2S test is the comparison of two models, which are fitted to the treated and untreated expression profiles for each probe. The first model fits a single Gaussian process to the time series data for both treatments, the second fits two independent Gaussian processes to the data. These two models are compared by calculating the noise within the fit and subtracting the two values to give the Bayes Factor (BF) score. This results in a ranked list, upon which a threshold can be decided to determine differential expression. A transcript is considered differentially expressed if the two independent Gaussian processes fit the data better than a single Gaussian process. The BF score indicates how strongly data support one theory, the higher the BF the greater the difference between the two models and the more likely a transcript is to be differentially expressed. The GP2S model fitting was carried out by Dr Christopher Penfold (University of Cambridge).

Transcripts which were considered differentially expressed were returned by GP2S and a cut off for the BF was determined by visual inspection of the output. A cut off of $BF > 14$ was used as the threshold, as this was the threshold at which no false positives were obvious in a visual inspection of a selection of 1,000 graphs with a BF score between $BF = 10 - 20$. Using this threshold, 13,638 transcripts were identified as differentially expressed. This list of differentially expressed transcripts mapped back to 8,918 Bo IDs and contained 1,925 transcripts which did not map to a Bo gene ID. These may include GD33DH specific genes, genes from the Cg-DFFS or genes not yet identified in the TO1000 sequence analysis.

Figure 4.5 shows GP2S output of the same exemplar transcripts that are illustrated in the ‘Treatment x Time’ and ‘Treatment’ categories in Figure 4.4 - *ERD15* and a MATE efflux protein transcript. In both cases it can be seen that two Gaussian processes provide a better fit to the data than a single Gaussian process, giving a BF score of $BF = 16.5$ for the *ERD15* transcript (Fig. 4.5a) and $BF = 44.7$ for the MATE efflux protein transcript (Fig. 4.5b).



(a) Gaussian Process Two Sample test on time series expression data of Bo4g190900.1, an *ERD15* transcript



(b) Gaussian Process Two Sample test on time series expression data of Bo2g047740.1, a MATE efflux protein transcript

Figure 4.5: Gaussian Process Two Sample test on time series expression data

A single Gaussian Process (GP, left) and two independent GPs (right) are fitted to the transcript expression profiles. The likelihood of each fit is given and the difference between the two likelihoods, the Bayes Factor is given. A Bayes Factor score of $BF=14$ or over indicated differential expression. Green is treatment, red is control.

Selecting differentially expressed transcripts

The analysis described above resulted in 13,638 transcripts differentially expressed according to GP2S analysis and 94,232 transcripts differentially expressed according to the *F*-Test ‘Treatment’ term analysis. Given the large number of differentially expressed transcripts resulting from the *F*-Test ‘Treatment’ term it was decided to include only transcripts which overlapped with the GP2S output in the final list of differentially expressed transcripts. The intersection of the two analyses resulted in a robust set of 11,754 transcripts that were identified as differentially expressed in both the GP2S test and the *F*-Tests. These transcripts mapped back to 7,141 Bo gene IDs (12% of the transcriptome), of which 737 were transcription factors. There were 1,333 probes which did not map back to a Bo gene ID thus representing potential novel transcripts with a role in salt shock. Differential expression was divided into either up-regulated or down-regulated in response to salt shock by calculating the difference between the expression mean of the salt treated and control samples for each transcript. If the difference was >0 , then the transcript was considered up-regulated, if <0 , then the transcript was considered down-regulated. The list of differentially expressed transcripts can be found in additional datafile3 (Appendix I).

Top 20 differentially expressed transcripts

Analysis of the potential functions of the most differentially expressed transcripts can give valuable insight into the most significant changes occurring in the transcriptome in response to salt shock in GD33DH and represents the ability of the transcriptome to rapidly adapt to a changing environment. The top 20 most differentially expressed transcripts, along with mapped Bo gene IDs and Arabidopsis orthologs annotations are shown in Figure 4.6. Within this list are many interesting expression profiles.

The most differentially expressed transcript identified by GP2S encodes a non-specific Lipid Transfer Protein (nsLTP), which shows enhanced expression from 2 htp (Fig. 4.6). nsLipid Transfer Proteins are found in abundance throughout the plant kingdom, with 63 putative LTPs identified in *B. rapa* (Li et al., 2014b). Expression of *nsLTPs* responds to ABA via cis-regulatory regions in the promoter (Tapia et al., 2013; Yubero-Serrano et al., 2003). The proteins play key roles in the stabilisation and organisation of the cell membrane (Boutrot et al., 2008) and respond to biotic stress and environmental stresses (Won et al., 2003; Yubero-Serrano et al., 2003). nsLTPs are also involved in biosynthesis of wax, which forms a protective barrier on the leaf surface protecting the leaf from further desiccation by reducing transpiration. This trait is increased in response to drought stress and is associated with increased levels of *LTP* (Tapia et al., 2013). The cell wall is severely disrupted following salt shock, caused by changes in osmotic potential, affecting the lipid





















Transcript	TO1000 ID	ATG	T01000 Descriptions	GP2S Bayes Factor	TOFDE	Profile
comp44278_c0_seq5	Bo3g023690.1	-	Non-specific lipid-transfer protein	104.1252	0h	
TCONS_00003597	Bo1g047810.1	AT4G26965	NADH:ubiquinone oxidoreductase subunit-17.2	99.8957	0h	
TCONS_00069445	-	-	-	90.2616	2h	
TCONS_00142551	Bo8g049770.1	AT4G17870	Absciscic acid receptor PYR1	89.0739	2h	
TCONS_00059012	Bo3g107210.1	AT5G62730	Major facilitator superfamily protein	86.307	2h	
TCONS_00086657	Bo5g013650.1	-	S-receptor kinase	85.0889	0h	
comp60489_c0_seq3	Bo3g165670.1	AT4G22920	Senescence-inducible chloroplast stay-green protein	84.5498	2h	
TCONS_00171561	Bo9g031100.1	AT1G64660	Cystathionine gamma-lyase	84.4843	2h	
TCONS_00167269	Bo9g164320.1	AT5G15250	ATP-dependent zinc metalloprotease FtsH	82.3784	18h	
TCONS_00036585	-	-	-	79.938	2h	
TCONS_00135780	Bo7g087520.1	AT3G29575	Ninja-family protein	78.9513	2h	
TCONS_00160887	Bo9g027620.1	-	Threonine synthase, putative	78.8117	2h	
TCONS_00162373	Bo9g059510.1	AT5G44020	Acid phosphatase 1, putative	78.8013	0h	
TCONS_00156706	Bo8g105870.1	AT1G15810	30S ribosomal protein S15	78.2604	2h	
comp60096_c2_seq1	Bo6g064670.1	AT5G40650	NAC domain containing protein	77.5407	2h	
TCONS_00160888	Bo9g027620.1	-	Threonine synthase, putative	75.9698	2h	
TCONS_00170786	Bo9g022010.1	AT5G66320	GATA transcription factor	75.9105	2h	
TCONS_00097024	Bo5g022200.1	AT1G16850	conserved hypothetical protein	75.5094	2h	
TCONS_00034484	Bo2g148140.1	AT3G29575	Ninja-family protein	75.3723	2h	
TCONS_00075777	Bo4g027570.1	AT2G38310	Absciscic acid receptor PYR1	74.9762	2h	

Figure 4.6: Top 20 differentially expressed transcripts

The top 20 differentially expressed transcripts in *B. oleracea* GD33DH in response to salt shock, their Bo gene ID and Arabidopsis ortholog, Bayes Factor score and expression profiles. The red represents the salt-treated expression profiles and blue is the control. The shaded areas represent the minimum and maximum values. TOFDE is Time Of First Differential Expression.

membrane layer thus the rapid up-regulation of nsLTP could be an attempt to repair damage caused to the cell membrane.

Transcripts mapping to two NINJA-family proteins were highly up-regulated (encoded by Bo7g087520.1 and Bo2g148140.1) in the course of the experiment. Both transcripts map to the same Arabidopsis gene (AT3G29575) which encodes *ABI five binding protein 3 (AFP3)* a gene that has three orthologs in *B. oleracea* (Parkin et al., 2014). Whilst the NINJA protein negatively regulates the expression of jasmonate related genes, the closely related AFP family of proteins interacts with ABI5 and TPL proteins to regulate ABA-related gene expression (Pauwels et al., 2010). That two of these orthologs are highly up-regulated in the experiment indicates that the regulation of the ABA response pathway in salt-treated GD33DH may be via this pathway.

Of other up-regulated transcripts, a transcript mapping to the *SENESCENCE INDUCIBLE STAY-GREEN (SGR)* protein (encoded by Bo3g16570.1) was highly up-regulated. The protein encoded by this gene is ABA-responsive (Delmas et al., 2013) and regulates chlorophyll degradation. Mutations in *SGR* orthologs cause a stay-green phenotype as the senescence process is delayed (Ren et al., 2007). In abiotic stress conditions, *SGR* has been shown to promote stress induced leaf yellowing during vegetative growth in Arabidopsis (Sakuraba et al., 2014b). This suggests a similar role in GD33DH, where senescence maybe induced in fully expanded leaves following salt shock.

Of the down-regulated transcripts, two transcripts mapping to an ABA receptor protein *PYR1* (encoded by Bo8g049770.1 and Bo4g027570.1) are present in the list and are down-regulated in both instances. That these transcripts map to different Arabidopsis genes suggests that these are not a set of triplicate genes as seen above. A similar expression pattern was reported in Arabidopsis after 24h of salt and osmotic stress, in which the PYR1 receptor protein was down-regulated however its target proteins SnRK2s, PP2Cs, ABI1 and ABI2 were up-regulated. It has been suggested that increasing the PP2Cs:PYR/PYLs ratio is important for the activation of downstream ABA responsive gene expression under abiotic stress (Chan, 2012). Though this is somewhat counter-intuitive, the results from this experiment would support this hypothesis.

A final down-regulated transcript to be discussed, maps to Bo8g105870.1 and encodes a 30S ribosomal protein S15. This transcript shows immediate loss of expression following salt shock, whilst in the control expression appears to be diurnal. Decrease in expression of this transcript highlights the effect that salt shock has on ribosomal proteins and hence the translation of novel proteins. Indeed, a decrease in protein translation is seen following water stress conditions such as drought (Huang et al., 2008) and salt stress (Omidbakhshfard et al., 2012).

4.4.2 Determining time of differential expression

In order to take full advantage of the temporal nature of the dataset, and to be able to provide time specific biological context to the transcripts, the expression profiles of each transcript were sub-divided based on the time at which the salt-treated profile becomes differentially expressed from the control profile using the GP2S Time Local method (Stegle et al., 2010), as carried by Dr Christopher Penfold (University of Cambridge).

The treatment and control expression profiles of the 11,754 significantly differently expressed transcripts were used in the analysis. The algorithm determines whether the expression profile of each transcript can be best explained using one Gaussian process or two Gaussian processes at each two hour time point. Here, a Gaussian process is referred to as an ‘expert’. If two experts are preferable to one, then that iteration is given a score of 1, else a score of -1 is given. Over 50 samplings (the Gibbs sampler) an average Z-indicator score is generated, as shown in Figure 4.7. If a transcript has an average Z-indicator score over 0, it is considered differentially expressed at that time point. This was used to establish at what time point the expression of a transcript is turned on or off during the time series experiment.

Figure 4.7 shows some examples of Z indicator profiles with their corresponding expression profiles for comparison. Some of the differentially expressed transcripts showed a gradual increase in expression throughout the experiment (Fig. 4.7a), whilst others showed a rapid change in expression (Fig. 4.7b). Differential expression was not necessarily maintained for the entire time series, examples of early differential expression in which diurnal expression is lost, followed by a return to control expression levels (Fig. 4.7c) and transcripts with ‘on then off’ induction of expression (Fig. 4.7d) were also seen.

Figure 4.8a shows the time at which transcripts first become differentially expressed i.e. the fitted experts diverge significantly from each other (up-regulated in light grey, down-regulated in dark grey). As many gene expression changes occur rapidly in response to salt shock, the expression of a collection of transcripts is already significantly differentially expressed by 2 hpt, therefore the experts diverge in the 0 hpt time frame. Based on this analysis it appears that there are two key time points upon which genes first become differentially expressed - 2 hpt and 18 hpt (Phase 1 and Phase 2). Figure 4.8b shows the total number of transcripts differentially expressed at each time point. Again, it appears there are two phases in which gene expression is grouped. Phase 1 - lasting between 0 and 16 hpt in which number of transcripts differentially expressed ranged from 750 - 6,000, the majority of which were down-regulated. Phase 2 - at 18 - 26 hpt a switch occurs and the number of differentially expressed genes rises to between 7,500 - 10,000, with peak expression occurring 24 hpt. It is possible that Phase 1 and Phase 2 correlate to the osmotic phase and ionic phase described by Munns and Tester (2008), therefore subsequent analysis focuses on the biological functions of transcripts that are first differentially expressed at

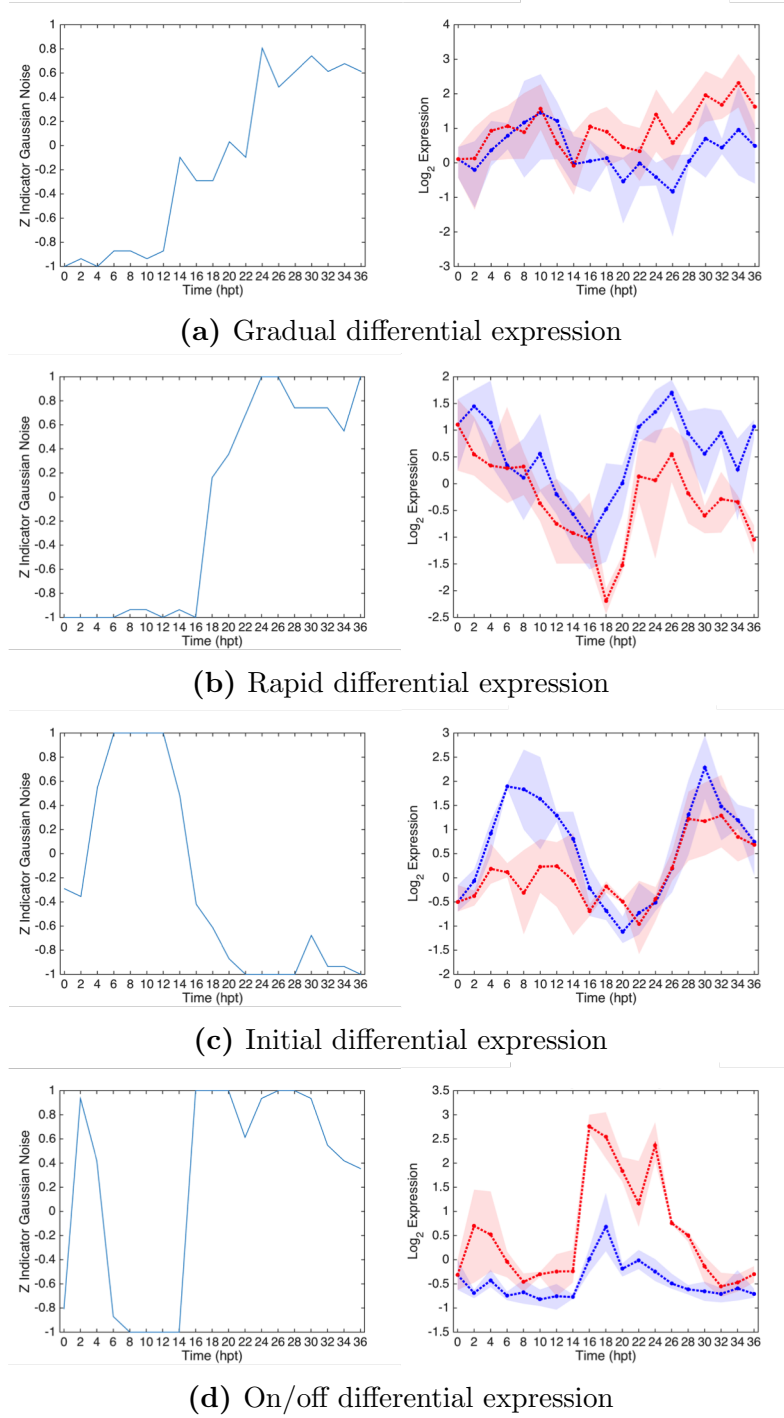


Figure 4.7: Z-indicator plots and corresponding transcript expression profile

Z-indicator plot and corresponding expression profile of a transcript which shows (a) a gradually increased expression compared to control from 22 hpt; (b) a transcript with a rapid switch to differential expression at 18 hpt; (c) a transcript with lost diurnal expression in first 20 hpt; (d) rapid burst of differential expression at 2 hpt, followed by a more sustained up-regulation after 16 hpt.

each time point.

The detailed analysis of individual transcript expression profiles has allowed the grouping of transcripts according to their time of differential expression. This allows the identification of co-ordinated, time specific changes in gene expression following salt shock and indicates early decrease of specific processes such as metabolic activity during the plant response to salt shock.

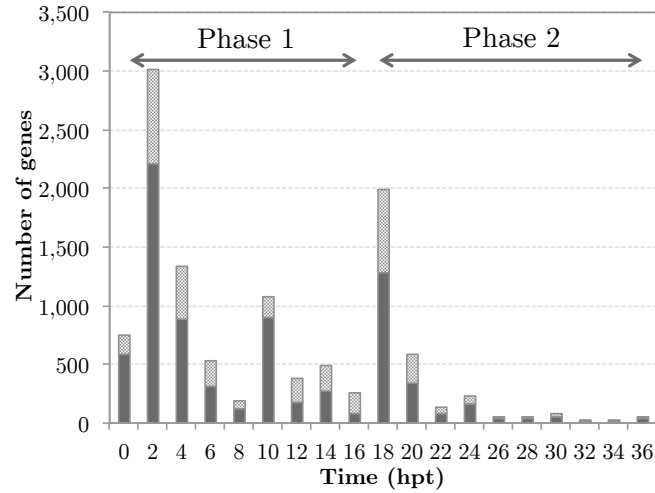
4.4.3 Establishing a chronology of the early salt shock response in *B. oleracea* GD33DH

The temporal information gained from the Time Local GP2S analysis was used to carry out GO analysis of Bo gene IDs mapping to transcripts at the time point in which they were first differentially expressed (Fig. 4.8a). This was carried out to determine whether there was a functional chronology to the salt shock response and whether different transcript functions could reflect the two different phases of gene expression seen in Figure 4.8. The results can be seen in Figure 4.9, which goes up to 24 hpt, after which too few transcripts are first differentially expressed to be able to carry out the analysis (Fig. 4.8a). The GO terms are annotated in red to indicate up-regulated transcripts and green to indicate down-regulated transcripts.

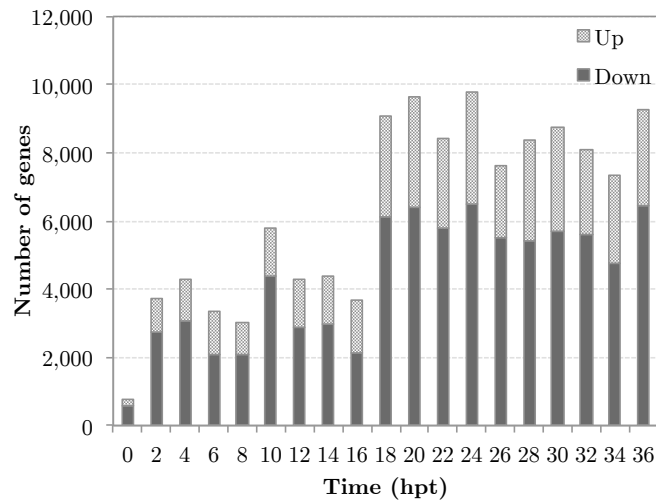
Early up-regulated transcripts have roles in the response to osmotic stress, initiation of stress signalling pathways, ion homeostasis (including osmolyte biosynthesis and ion transport). The ABA response was underway within 2 hpt indicating rapid signalling mechanisms are involved following salt shock. ABA has important functions under osmotic stress, acting as a regulator of stomatal aperture and movement, as indicated at 2 hpt. ABA does not function in isolation, there is cross talk between other signalling pathways as indicated by the presence of over-represented GO terms relating to JA and SA (4 hpt) forming a signalling network which can be finely tuned as needed.

Ion homeostasis was rapidly up-regulated at 2 hpt in which ion transporters were recruited to maintain high a $K^+ : Na^+$ ratio by keeping the concentration of Na^+ in the cytoplasm as low as possible through sequestration of excess Na^+ ions in the vacuole. The biosynthesis of proline was also up-regulated at 2 hpt, an important osmolyte with roles in protecting cells from the further damages caused by excess Na^+ ions. In the later stages of the time series, after 12 hpt, various methods of transport were up-regulated including vacuolar transport, vesicle mediated transport (18 hpt) and intracellular transport (20 hpt) suggesting that GD33DH is sequestering excess ions in the vacuole.

Up-regulation of peroxisome related transcripts was seen between 16 - 18 hpt. Peroxisomes are small vesicles containing a number of enzymes involved in processes such as fatty acid β oxidation and the oxidation of glycolate by glycolate oxidase in photorespiration



(a) Time of first differential expression of transcripts



(b) Differentially expressed transcripts at each time point

Figure 4.8: Differentially expressed transcripts at each time point

(a) The time at which each transcript first becomes differentially expressed. (b) Differentially expressed transcripts at each time interval in response to salt shock in GD33DH. There is a natural split at 18 hpt in which the number of differentially expressed transcripts increases drastically. This split has been term ‘Phase 1’ for 0 - 16 hpt and ‘Phase 2’ for 18 - 36 hpt.

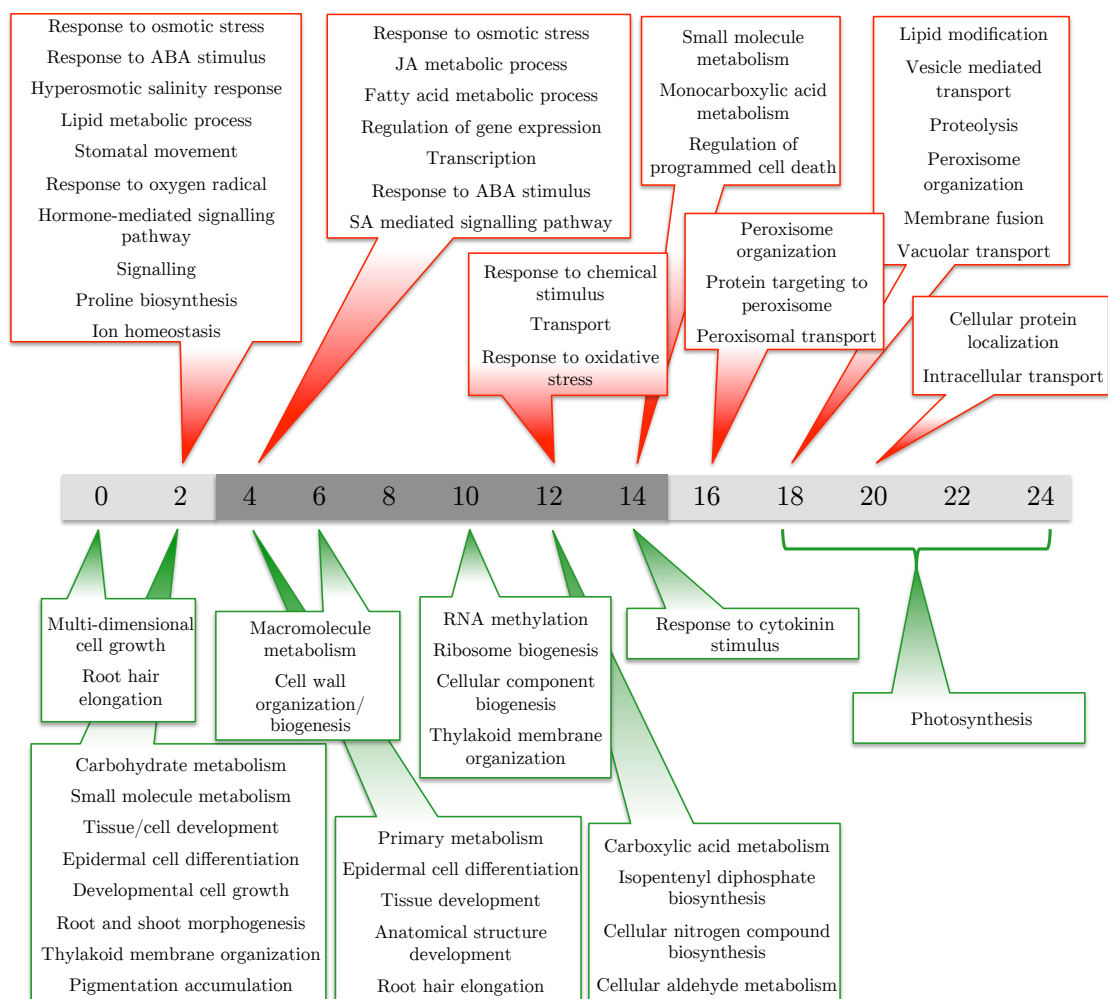


Figure 4.9: Selected over-represented GO terms over 24h of transcripts differentially expressed after salt shock in *B. oleracea* GD33DH

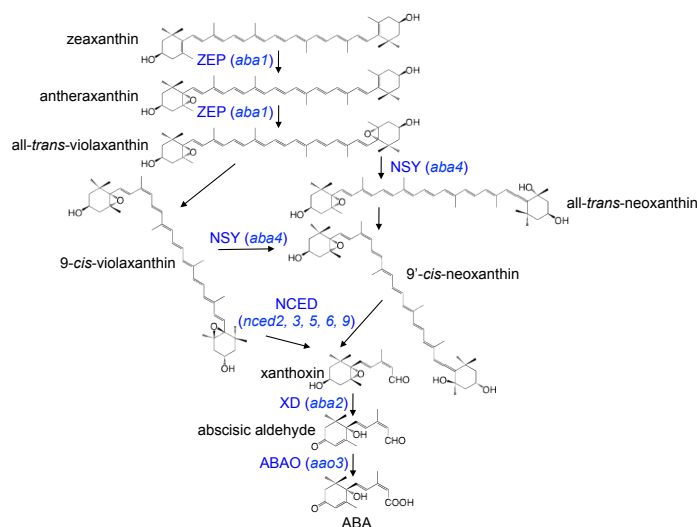
Enriched GO terms were identified using BiNGO (Maere et al., 2005b) in groups of transcripts that first show significant up-regulation or down-regulation in response to salt shock over the time series up to 24 hpt. Red boxes contain GO terms from up-regulated genes and green boxes contain GO terms from down-regulated genes. GO terms are ranked in order of significance (adjusted P-value) (Benjamini and Hochberg, 1995).

occurs in peroxisomes. Peroxisomes produce large levels of H_2O_2 and may have roles in signalling through ROS production (reviewed in Sharma et al., 2012). An increase in the number of peroxisomes is seen under salt stress conditions in *Arabidopsis* but has no overall effect on the tolerance (Mitsuya et al., 2010). Finally, lipid modification and proteolysis are up-regulated at 18 hpt suggesting that the damage to the cell membrane and protein content of the cell is in the process of repair suggesting that the initial ion excess is under control by this time point.

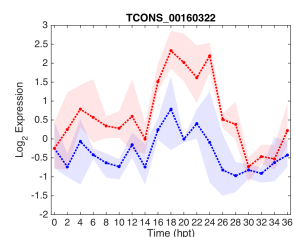
In contrast, early down-regulated transcripts have roles in growth and development of the root and shoot, progressing to the down-regulation of metabolism and biosynthesis. Under salt stress, particularly during the osmotic phase of the response, growth of the root and shoot is suspended to reduce leaf area of the plant thus reducing water loss by transpiration through the stomata (Munns and Tester, 2008). Root architecture, particularly the formation of lateral roots, is affected through withdrawal of auxin and ABA sensitivity and plants can alter their root growth upon detection of high saline (Brady et al., 2003; De Smet et al., 2003). Many transcripts with GO terms relating to cell growth and development were found to be down-regulated between 0 - 4 hpt indicating that the down-regulation of growth and development is in action. In addition, at 14 hpt, transcripts relating to cytokinin stimulus are down-regulated. Down-regulation of cytokinin activity has previously been associated with reductions in growth (Fig. 4.9).

Transcripts with associations to biosynthesis are down-regulated from 10 hpt, including ribosome biogenesis and cellular component biogenesis. It has been previously shown that high levels of Na^+ can affect the function of ribosomes, decreasing the capacity to synthesise new proteins. A decrease in ribosome biogenesis has been correlated with decreased growth (Galvan-Ampudia and Testerink, 2011), therefore this down-regulation of energetically expensive biosynthesis and ribosome biogenesis could represent an adaptive mechanism, to conserve energy for other processes such as ion transport and biosynthesis of osmolytes such as proline.

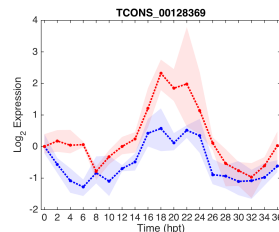
GO terms involved in the later stages of the salt shock response are primarily related to a reduction in photosynthesis (18 - 24 hpt). It must be remembered that these GO terms relating to the time at which transcripts associated with them are first switched on, indicating that photosynthesis is primarily down-regulated during the later stages of the salt shock response. It has been shown that increased levels of ABA leads to a decrease in stomatal opening and reduced CO_2 availability, as well as a decrease in photosynthetic related gene expression (reviewed in Chaves et al., 2008).



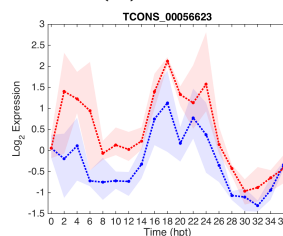
(a) ABA biosynthesis pathway



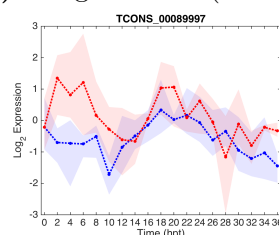
(b) Bo9g020440.1 (*ABA1*)



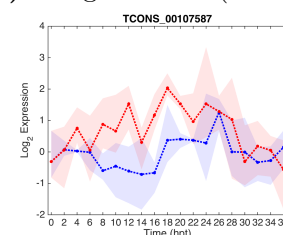
(c) Bo7g105930.1 (*NCED2*)



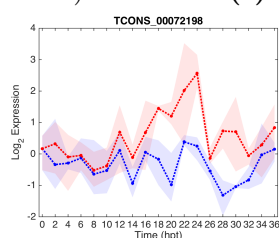
(d) Bo3g066190.1 (*NCED3*)



(e) Bo5g062360.1 (*NCED5*)



(f) Bo6g028000.1 (*ABA2*)



(g) Bo4g161370.1 (*AAO3*)

Figure 4.10: ABA biosynthesis and signalling

(a) An overview of the ABA biosynthesis pathway (adapted from hormones.psc.riken.jp). In blue are the enzymes and respective gene names are given in parentheses. Expression profile plots of transcripts mapping to ABA biosynthesis genes (b) *ABA1*; (c) *NCED2*; (d) *NCED3*; (e) *NCED5*; (f) *ABA2*; (g) *AAO3*. Log₂ expression on the *y*-axis and time on the *x*-axis. Red corresponds to the salt-treated expression profile, blue to the control and shaded areas indicate minimum/maximum values.

4.4.4 Transcripts involved in ABA biosynthesis are up-regulated following salt shock

It has been well established that ABA is the primary hormone in regulating the plant response to abiotic stress and that it acts in concert with ethylene, auxin, JA, cytokinins, brassinosteroids and SA to regulate gene expression under stress conditions (Chan, 2012; Shinozaki et al., 2003; Tran et al., 2010; Zhang et al., 2006). Expression of several transcripts mapping to Bo gene IDs that encode proteins involved in the ABA biosynthesis pathway were up-regulated in response to salt shock in GD33DH (Figure 4.10). ABA biosynthesis occurs by several enzymatic reactions (Fig. 4.10a), where ABA1 (encoding the zeaxanthin epoxidase (ZEP) enzyme, Bo9g020440.1) epoxidates zeaxanthin, the products of which are then converted by 9-cis-epoxycarotenoid dioxygenase (NCED2, NCED3 and NCED5, encoded by Bo7g105930.1, Bo3g066190.1 and Bo5g062360.1, respectively) to form xanthoxin. The short-chain xanthoxin dehydrogenase (XD) that is encoded for by ABA2 (Bo6g028000.1) catalyses the conversion of xanthoxin to abscisic aldehyde which is oxidised into ABA by Arabidopsis aldehyde oxidase 3 (AAO3, Bo4g161370.1) (Barrero et al., 2006; Gonzalez-Guzman et al., 2002; Nambara and Marion-Poll, 2005). These ABA biosynthesis genes were significantly up-regulated by 2 hpt with the exception of *ABA2* (4 hpt) and *AAO3* (14 hpt), indicating an early role for the hormone in response to salt shock.

That this entire pathway, with the exception of *ABA4*, is up-regulated in GD33DH in this experiment indicates the importance of this hormone signalling pathway during the salt shock response. Many stress response genes are mediated by ABA, so biosynthesis of the phytohormone in response to the stress is crucial to elicit the correct response.

4.4.5 Modelling ABA signalling following salt shock

ABA produced in response to abiotic stress activates the PYR group of ABA receptors, which then inhibit PP2Cs leading to the activation of SnRK2s through autophosphorylation (Ma et al., 2009; Park et al., 2009; Santiago et al., 2009). SnRKs are involved in mediating the ABA response by phosphorylating downstream targets, such as ABF2, ABF4 and ABF3 TFs which leads to the activation of many down-stream ABA regulated genes (Yoshida et al., 2010). An advantage of high-resolution time series expression data is that it can be used in biological network inference, to predicted the topology of a gene regulatory network. It was decided to model the expression patterns of PYRs, PP2Cs and SnRK2s, to determine how the perception of ABA through the PYR/PP2C/SnRK2 pathway is regulated under salt shock conditions.

The Causal Structure Inference (CSI) algorithm (Penfold et al., 2012, 2015; Penfold and Wild, 2011) was used to infer a complex gene regulatory network involved in fine

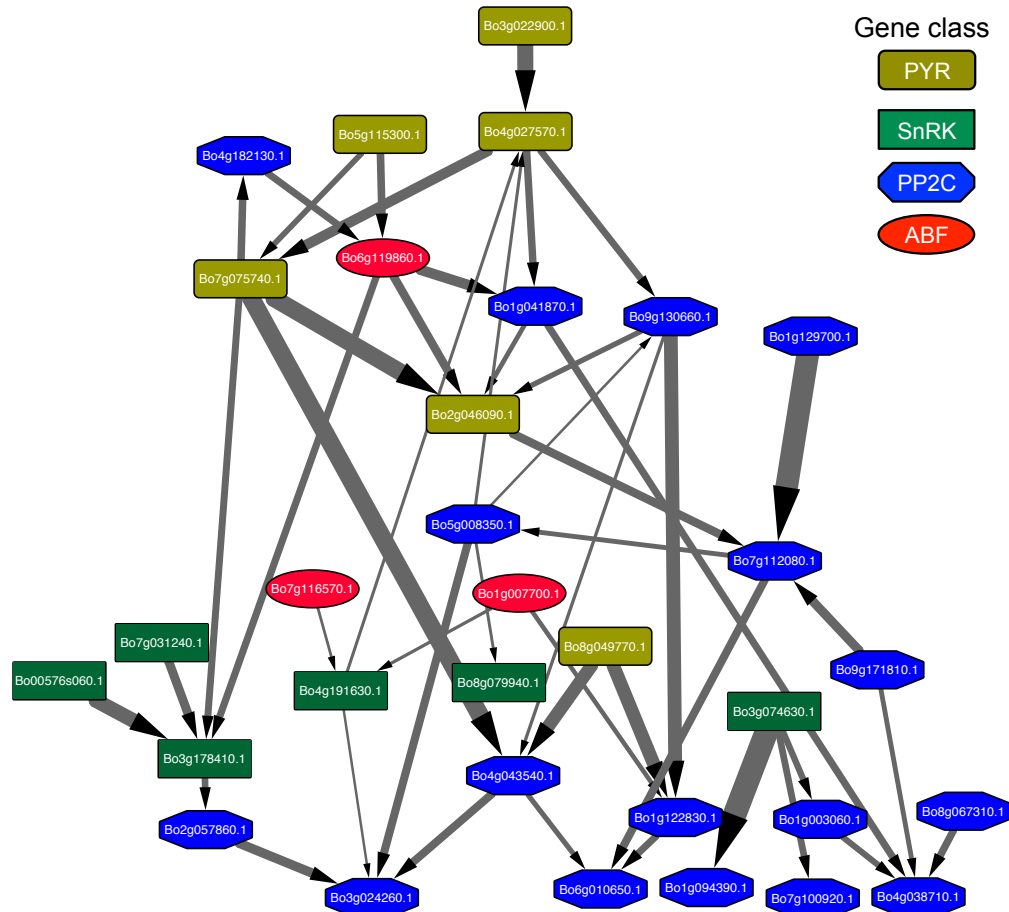


Figure 4.11: Casual Structure Identification network inference of ABA signalling components

Network inference of ABA signalling components inferred using the CSI algorithm (Penfold et al., 2012, 2015). Each node represents a differentially expressed transcript relating to Bo genes with annotations relating to PYR, PP2C, SnRK or ABF. The edge indicates the direction of regulation, but does not indicate activation or repression.

tuning the regulation of ABA perception and hence downstream expression of ABA-related genes. In total, 32 transcripts mapping to Bo genes with annotations of PYR, PP2C, SnRK and ABFs were included in the final CSI network model (Figure 4.11). The model showed hierarchy, though not in a linear fashion (as described above), but rather as an interconnected network suggesting complex cross talk between components within the ABA receptor pathway. It was clear from the model that the PYR genes sit at the top of the model and control the downstream interactions between SnRKs, PP2Cs and ABFs. The ABF genes do not sit at the bottom of the network, regulating expression of ABA responsive genes with an AREB binding domain, as described above. However, the CSI algorithm does not give information on the type of regulation, for instance activation or repression of downstream target genes, which is important in the ABA signalling model (Ma et al., 2009; Park et al., 2009; Santiago et al., 2009).

The role of PYR/PYL/RCAR receptors in regulation and stomatal aperture and ABA responsive gene expression was established through analysis of a sextuple mutant with 6 impaired PYR/PYL receptors suggesting an important role for this group of proteins in ABA signalling, stomatal aperture, germination and growth (Gonzalez-Guzman et al., 2012). Leveraging of ABA signalling through manipulation of the PYR/PYL/RCAR receptors has shown increased water use efficiency in Arabidopsis (Yang et al., 2016). Manipulation of the PYR/PYL/RCAR receptors in tomato has shown potential for enhanced drought tolerance (Gonzalez-Guzman et al., 2014).

The model in GD33DH suggests that the ABA signalling network is controlled by multiple PYR receptors. There is extensive cross talk and complex regulation between different components of the ABA signalling pathway in order to fine tune the response to salt shock in GD33DH (Figure 4.11).

4.4.6 Hormone cross talk in the regulation of the salt shock response

As a result of this chronological analysis, it was decided to investigate hormone related transcripts over the time series and look in more detail at ion transporters and their roles in the response to salt shock in GD33DH. Analysis of the time of first differential expression of differentially expressed transcripts mapping to Bo gene IDs with GO terms for ‘response to abscisic acid’, ‘response to jasmonic acid’ and ‘response to ethylene’ revealed a large amount of potential cross talk between the hormone signalling pathways in response to salt shock (Fig. 4.12). Transcripts associated with ABA signalling were more abundant than transcripts associated with JA and ethylene, indicating the importance of this hormone in response to salt shock in GD33DH. Other hormones related genes, JA and ethylene were still altered in expression, suggesting crosstalk between the signalling pathways in

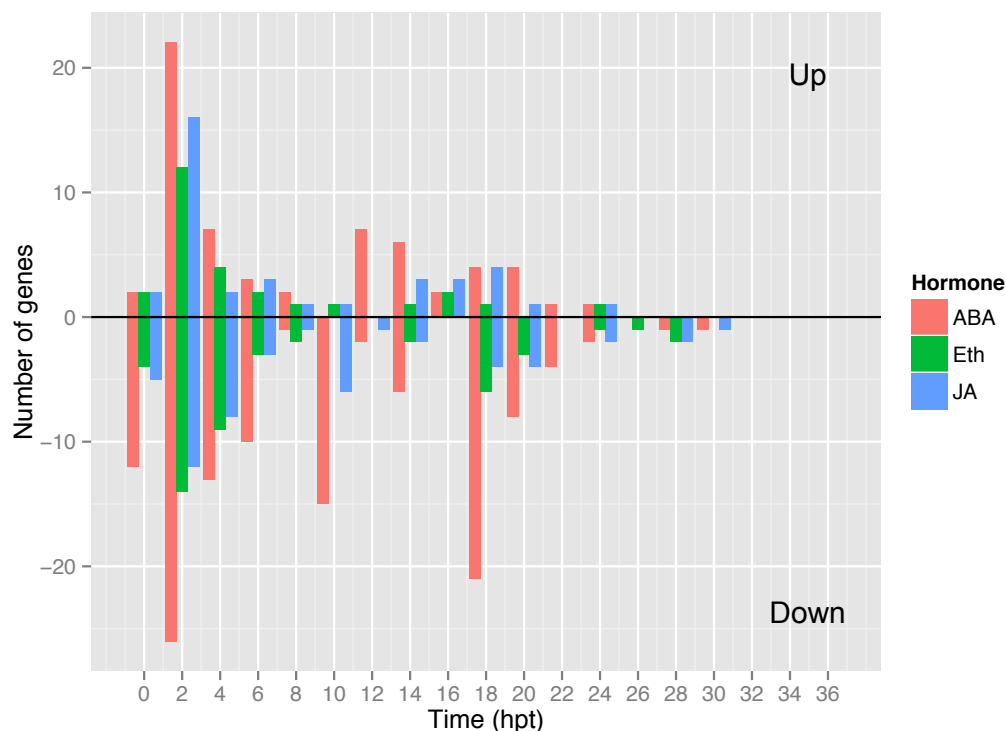


Figure 4.12: Time of first differential expression on selected groups of transcripts

Bo gene IDs mapping to differentially expressed transcripts associated with the GO terms ‘response to abscisic acid’ (184/801 gene, labelled ‘ABA’), ‘response to jasmonic acid’ (88/488 genes, labelled ‘JA’) and ‘response to ethylene’ (74/471, labelled ‘Eth’) were examined using the Time Local GP2S tool for time of first differential expression. Transcripts not mapping to a Bo gene ID were discarded from analysis.

response to salt shock, as has been seen in the literature in abiotic stress responses. The time point at which most hormone related transcripts were first differentially expressed was 2 hpt, suggesting hormone signalling is occurring in the early stages of salt shock, as has also been shown extensively in the literature (Anderson et al., 2004; Chan, 2012; Seki et al., 2002; Shinozaki et al., 2003; Yoshida et al., 2014b).

4.4.7 Differentially expressed ion transporters

A collection of ion transporter proteins were differentially expressed in the experiment and are shown in Table 4.3. This list includes ion transporters such as MATE efflux family proteins, ABC transporter proteins, aquaporins and major facilitator superfamily proteins. Most ion transporters that are involved in removing Na^+ are expressed in the root and have roles in preventing Na^+ uptake and xylem loading, whilst increasing K^+ uptake to maintain a high $\text{Na}^+:\text{K}^+$ ratio. Inevitably, as seen in Table 5.7 Na^+ levels rise in the shoot, alongside a drastic rise in K^+ , with the maintenance of a control $\text{Na}^+:\text{K}^+$ ratio by 36 hpt.

Typical salt stress transport proteins such as *SOS1*, *HKT1* and *HKT2* were not found

Transporters	Number	Direction	Arabidopsis orthologs
ABC transporter	33	8 up, 25 down	<i>AtABCA1</i> , <i>AtABCA2</i> , <i>AtABCB19</i> , <i>AtABCG14</i>
Aquaporin	17	4 up, 13 down	<i>AtGAMMA-TIP</i> , <i>AtRD28</i> , <i>AtPIP1B</i>
Cation transporters	2	2 down	<i>AtCCC1</i> , <i>AtCLC-A</i>
CNGC	7	3 up, 4 down	-
Glutamate receptor	6	2 up, 4 down	<i>GLR2.7</i> , <i>GLR2.9</i> , <i>GLR3.3</i> , <i>GLR3.5</i> , <i>GLR5</i>
Glutathione-regulated potassium-efflux system protein	2	2 down	<i>AtKEA3</i>
H-ATPase	2	2 down	<i>AtHA1</i> , <i>AtHA2</i>
Major facilitator protein	18	7 up, 11 down	-
Sodium/hydrogen exchanger	5	2 up, 3 down	<i>AtNHX1</i>
MATE	34	17 up, 17 down	-
Potassium channel	2	2 up	<i>AtKT2/3</i>
Potassium transporter	5	2 up, 3 down	<i>AtKUP6</i> , <i>AtKUP10</i> , <i>AtKUP11</i>
V-type proton ATPase	25	10 up, 15 down	<i>AtVHA-A1</i> , <i>AtVHA-A2</i>
Vacuolar cation/proton exchanger	1	1 down	<i>AtCAX1</i>
Voltage-gated potassium channel beta subunit	1	1 down	<i>AtKAB1</i>

Table 4.3: Ion transporter families

The number of differentially expressed transcripts mapping to key ion transporter families, the number up- and down-regulated and any noteworthy Arabidopsis orthologs.

to be differentially expressed in the shoot. These proteins mainly function in the root, at the site of Na^+ uptake and in the xylem parenchyma cells, to prevent xylem loading and transport to the shoot. Else, Na^+ ions can reach the shoot through the bypass flow mechanism in which flow through the apoplastic space occurs. Excess Na^+ and K^+ ions can be sequestered in the vacuole (both in the root and shoot) to prevent excess Na^+ ions causing ionic stress in the cell.

Many other classes of ion transporters that were differentially expressed in the time series experiment, interestingly the majority of which were down-regulated in the shoot under salt shock conditions suggesting different mechanisms of action for different families of ion transporters. For instance, the Glutathione-regulated potassium-efflux system genes e.g. *KEA3* ion transporters are found located on the thylakoid membrane, so down-regulation of this transporter may reflect a decrease in photosynthetic activity.

Many of the differentially expressed transporters were located on the vacuole membrane suggesting that Na^+ ions are stored in the vacuole to prevent Na^+ accumulation in the cytosol reaching toxic levels eg ABC transporters, H-ATPase, V-type proton ATPase, aquaporin, Vacuolar cation/proton exchanger, Sodium/hydrogen exchanger, MATE efflux protein.

There were a large number of K^+ transporters, such as those encoding orthologs to *KUP6*, *KUP10*, *KUP11* and *KT2/3* either up- or down-regulated in response to salt shock (Table 4.3). KT/KUP/HAK potassium transporters are on the vacuole membrane and may have a role in the efflux of K^+ (Grabov, 2007), suggesting that the proteins encoding these genes may be functioning to maintain the $\text{K}^+:\text{Na}^+$ ratio within the cell in an attempt to exclude additional Na^+ ions.

The differential expression in both directions of ion transporters suggests complex interplay between transporter proteins in order to maintain cellular homeostasis. The abundance of differentially expressed K^+ transporters may be responsible for maintaining the high K^+ , low Na^+ ratio within the cytosol, which is essential for salt tolerance. The high number of vacuolar ion transporters differentially expressed suggests compartmentalisation of intracellular Na^+ ions within the vacuole, to protect cells from cytotoxicity caused by high Na^+ content.

4.4.8 Differential expression of transcripts mapping to key genes with known involvement in abiotic stress responses

Expression levels of selected GD33DH transcripts, whose Arabidopsis orthologs have been shown to have a key role in the response to abiotic stress conditions, particularly dehydration stress are shown in Figure 4.13. This group includes some important genes which have been proposed to be involved in the abiotic stress response through the ABA, auxin or ethylene signalling pathways in Arabidopsis and other plant species. Transcript expression was compared to expression of Arabidopsis orthologs from the AtGenExpress dataset in salt stress and osmotic stress conditions using the Arabidopsis eFP Browser through bar.utoronto.ca (Winter et al., 2007).

GD33DH orthologs of several ABA inducible genes were found to be differentially expressed in this study, reaffirming the important role of this hormone in the response to salt shock. Of these *MYC2* (encoded by Bo5g086990.1; Fig. 4.13a) is an important bHLH TF involved in the cross talk between multiple stress response pathways. The MYC2 protein functions by binding to MYC recognition sites in the promoters of genes under its control. It has a well established role in the regulation of ABA inducible genes such as *RD22* (encoded by Bo9g011300.1; Fig. 4.13b) (Abe et al., 2003), of which the protein is a positive regulator of JA signalling (Pauwels et al., 2010). *RD22* expression has been used as a marker of ABA induced expression in drought conditions (Yamaguchi-Shinozaki and Shinozaki, 1993). In the *B. oleracea* GD33DH time series experiment, both *MYC2* and *RD22* gene expression levels were similar in both the treated sample and the control until 16 hpt, after which expression levels of both genes increased in the salt treated plants (Fig. 4.13a and b). The transcripts had a BF score of BF=29.2 and BF=15.8, respectively.

ATAF2 (encoded by Bo2g009250.1; Fig. 4.13c) is a member of the NAC transcription factor family that has been widely implicated in the biotic and abiotic stress responses (Ooka et al., 2003). The ATAF2 protein has been found to repress the expression of PR genes in biotic stress responses and is induced by dehydration independently of ABA in abiotic stress responses (Delessert et al., 2005). It is found as a protein partner to ATAF1 and maybe functionally redundant (Wu et al., 2009). The transcript had a BF

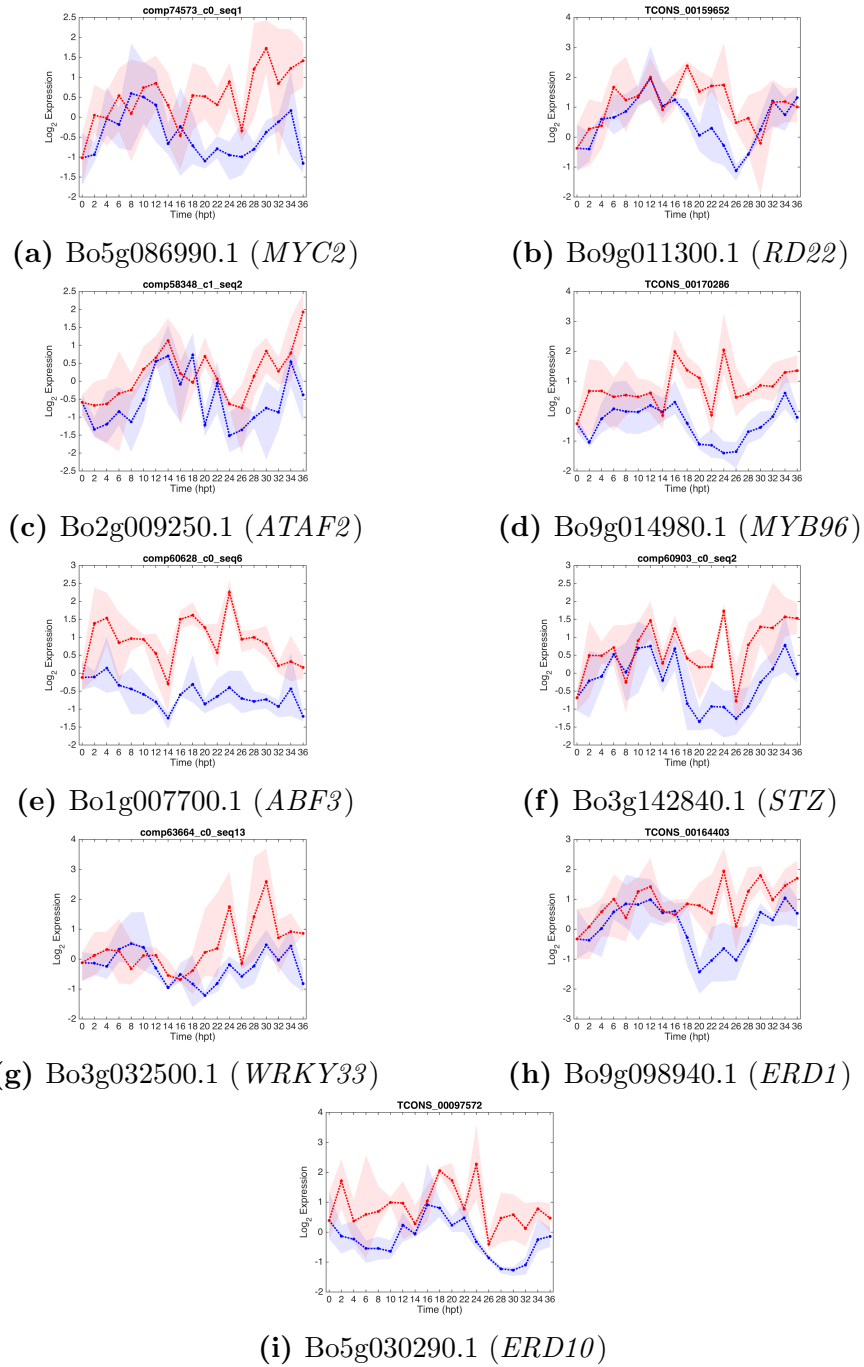


Figure 4.13: Expression profiles of differentially expressed *B. oleracea* GD33DH transcripts whose orthologs have previously reported functions in the abiotic stress response in other plant species

Plots of a selection of GD33DH differentially expressed salt shock genes, with the closest Arabidopsis ortholog. (a) Bo5g086990.1 (*MYC2*); (b) Bo9g011300.1 (*RD22*); (c) Bo2g009250.1 (*ATAF2*); (d) Bo9g014980.1 (*MYB96*); (e) Bo1g007700.1 (*ABF3*); (f) Bo3g142840.1 (*STZ*); (g) Bo3g032500.1 (*WRKY33*); (h) Bo9g098940.1 (*ERD1*) and (i) Bo5g030290.1 (*ERD10*). Log₂ expression on the *y*-axis and time on the *x*-axis. Red corresponds to the salt-treated expression profile, blue to the control and shaded areas indicate minimum/maximum values.

score of BF=14.0, which is at the bottom threshold level for differential expression. The up-regulation of this repressor protein could be an example of the plant lowering its basal immunity to redirect energies to rebalance following an increase in Na⁺ ions in the leaves.

MYB96 (encoded by Bo9g014980.1; Fig. 4.13d) has a role in stomatal movement, as well as being an important regulator in the cross talk between the ABA and auxin response pathway during lateral root development under water stress conditions (Seo et al., 2009). When the Arabidopsis ortholog was examined using the Arabidopsis eFP browser, it was shown that *MYB96* was differentially expressed mildly in leaf and strongly in the root at 3 hpt and levels were maintained up to 24 hpt, in both salt and osmotic stress treatments. In this experiment, accumulation of the MYB96 transcript occurred at 16 hpt in GD33DH following salt shock conditions. The protein likely plays an important role in stomatal movement and root development in response to salt shock in GD33DH.

Genes such as *ABF3* (encoded by Bo1g007700.1; Fig. 4.13e), along with *AREB1* and *AREB2* (plots not shown) encode proteins that have been found to be master regulators in ABRE-dependent ABA signalling during water stress conditions. They are bZIP transcription factors which work either as homodimers or heterodimers and require ABA for full activation of downstream gene expression (Yoshida et al., 2010). In this experiment, *ABF3* was instantly up-regulated and expression remained high for the duration of the experiment. It is likely the proteins that these transcripts encode play important roles in ABA-dependent gene expression, suggesting cross talk between stress response pathways in response to salt shock. Expression of the Arabidopsis ortholog in the eFP browser shows that the gene was differentially expressed in the leaf, and mildly in the root for both salt and osmotic stress between 0.5 and 6 hpt suggesting an early role for this gene in Arabidopsis.

SALT TOLERANCE ZINC FINGER (STZ) (encoded by Bo3g142840.1; Fig. 4.13f) is an abiotic marker gene whose protein has been implicated in salt and cold stress tolerance and is rapidly up-regulated under these conditions (Sakamoto et al., 2004; Seki et al., 2002; Teige et al., 2004). It is thought to have a role in repressing photosynthesis and carbohydrate metabolism and transgenic over-expressers show reduced growth (Maruyama et al., 2004). When the expression patterns of the Arabidopsis ortholog were examined in the Arabidopsis eFP browser, it was shown that the gene was differentially expressed in the roots in response to salt shock between 3 and 6 hpt. In the GD33DH experiment, the transcript steadily accumulated and became differentially expressed at around 16 hpt, suggesting that it plays a role in the repression of photosynthesis in response to salt shock.

WRKY33 (encoded by Bo3g142840.1; Fig. 4.13g) has been reported to play key roles in multiple stress responses including salt stress (along with WRKY25) (Jiang and Deyholos, 2008) and heat stress (Li et al., 2011). The downstream targets of WRKY33 include genes with important functions in responding to ROS for example peroxidases and

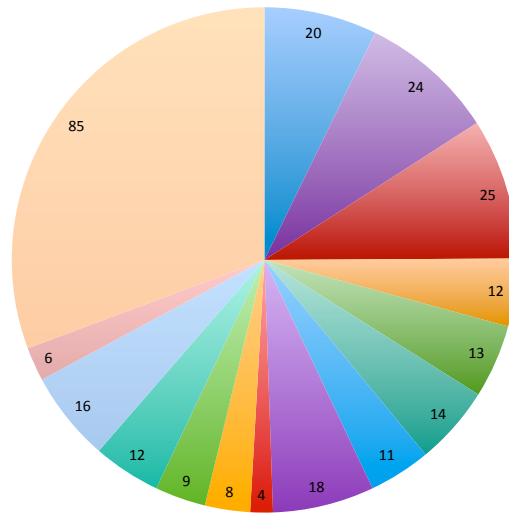
glutathione-S-transferases (Jiang and Deyholos, 2008) making it a key regulator in osmotic stress conditions. Viewing the expression patterns in the Arabidopsis eFP browser shows that the gene is differentially expressed in the roots at 3 hpt in response to salt treatment but not in the leaves. In this experiment, *WRKY33* became differentially expressed at around 18 hpt, where it accumulated above the control, suggesting additional roles in GD33DH, causing differential expression in the leaves and not in the leaves of Arabidopsis under similar experimental conditions.

Bo9g098940.1 and Bo5g030290.1 encode the EARLY RESPONSE TO DESICCATION proteins, ERD1 and ERD10, respectively; Fig. 4.13h, i). The *ERD* genes were rapidly activated upon drought stress in Arabidopsis (Kiyosue et al., 1994; Taji et al., 1999) and were highly up-regulated in the time series experiment, suggesting an important role in the response to salt shock in GD33DH. ERD1 encodes a chloroplast ATP-dependent protease (Soitamo et al., 2008), and is seen here to be up-regulated at 18 hpt after treatment. The Arabidopsis ortholog of this gene was differentially expressed in the leaves of osmotic stress conditions from 12 hpt. *ERD10* is a member of the late embryogenesis abundant protein (LEA) family that is up-regulated immediately, within 2 hpt. The expression pattern of this gene in Arabidopsis using the eFP browser shows that this gene is rapidly differentially expressed under both salt and osmotic stress conditions in both the leaf and root, though the effect is stronger in the leaf. The function of this gene family remains unclear, but roles have been proposed in the sequestration of ions (Bray, 1993) and a chaperone role protecting and refolding of proteins following water stress (Kovacs et al., 2008).

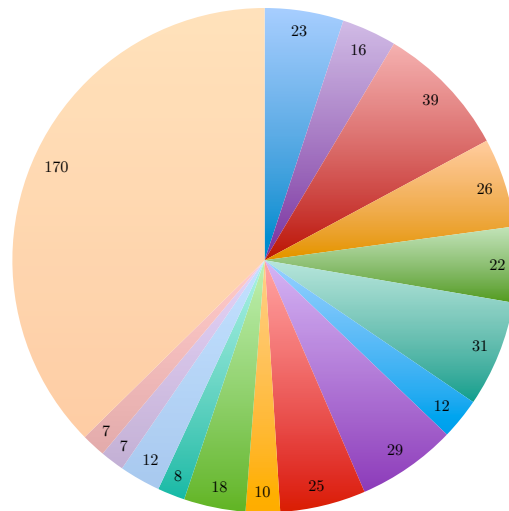
4.4.9 Differentially expressed transcription factor families

TFs are proteins that bind to DNA upstream of the coding region of a gene and regulate its expression by either recruiting or blocking the assembly of basal transcriptional machinery and of RNA polymerase II, which catalyses translation of DNA into mRNA. TFs bind to DNA binding domains, which are usually found upstream of genes. Multiple genes with related function can have the same DNA binding motifs and hence a single TF can regulate a multitude of genes. The TFs ability to regulate gene expression is dynamic, allowing rapid changes in the expression levels of the target genes depending on external stimuli. In order to assess the complex network of signalling pathways in GD33DH in response to salt shock, further analysis of TF families in particular was carried out.

A total of 737 differentially expressed transcripts encoding putative TFs were identified. Of particular abundance and interest were the bHLH, MYB and MYB-related, AP2-EREBP, bZIP, WRKY and NAC TF families which have been widely implicated in the response to abiotic stress conditions in genome-wide analyses of abiotic stress condition in various plant species (Hu et al., 2015; Liu et al., 2015a; Peng et al., 2014), and have been extensively reviewed (Baldoni et al., 2015; Chen et al., 2012; Llorca et al., 2014; Mizoi et al., 2012;



(a) Up-regulated TF families



(b) Down-regulated TF families

Figure 4.14: Differentially expressed TF families

(a) Up-regulated and (b) down-regulated differentially expressed TF families. The numbers in each segment represent the number of transcripts in the category, as per the key.

Nakashima et al., 2014; Puranik et al., 2012).

Transcription factors were identified by a querying the list of differentially expressed transcripts against the Plant Transcription Factor DataBase (Plant TFDB3.0) (Jin et al., 2013) and against the Arabidopsis TAIR10 assembly to obtain the closest ortholog, and was broken into up- and down-regulated, as summarised in Figure 4.14. A selection of transcripts with a named Arabidopsis ortholog will be discussed in the literature analysis of the results below, as shown in Appendix E.

AP2/EREBP transcription factor family

APETALA 2/ ethylene response element binding protein (AP2/EREBP) transcription factors are involved in stress acclimation by modulating cross talk between hormone signalling pathways (Dietz et al., 2010; Yang et al., 2005). They signal through the ethylene signalling pathway, which is often referred to as the ABA-independent stress responsive pathway. In *B. rapa* ssp. *pekinensis* 291 putative AP2/EREBP TF proteins were identified which could be further resolved into 15 groups - ARP2, ERF, RAV and Soloist (Song et al., 2013). Of the differentially expressed transcript list, 36 transcripts mapped to the AP2/EREBP TF family. The most well known genes in this TF family are the *DREB1* and *DREB2* TFs, belonging to the ERF group. Despite these genes being key in the salt response signalling through the ABA-independent pathway in Arabidopsis (Lata and Prasad, 2011; Oh et al., 2005), only *DREB2B* was differentially expressed in GD33DH during the first 36h of salt shock. Other AP2/EREBP genes were differentially expressed including the *ETHYLENE RESPONSE FACTOR 4* and *5* (*ERF4* and *ERF5*) genes. ERF4 is a repressor of expression and is capable of modulating both ethylene and ABA signalling (Yang et al., 2005) and ERF5 has been highly implicated in JA/Ethylene signalling in defence against pathogens such as *B. cinera* (Moffat et al., 2012) and in response to chitin (Son et al., 2011).

bHLH transcription factor family

The most abundant TF family in both up- and down-regulated differentially expressed transcripts was the bHLH TF family. In Arabidopsis, this group consists of a total of 162 bHLH genes which can be further split into 21 subfamilies (Toledo-Ortiz et al., 2003). In addition 167 bHLH genes have been identified in rice (Li et al., 2006). The functions of the bHLH TF cover a broad range of growth, developmental and maintenance processes that occur at all stages of the plant life cycle. Here, 62 transcripts mapping to bHLH TFs were identified as differentially expressed in the experiment. Down-regulated transcripts have roles in phytochrome signalling (PIF3 and PIF4), dark-induced senescence (PIF4 and PIF5) (Sakuraba et al., 2014a), and the response to far red light (controlled by

PIF3, PIF4 and PIF7) (Leivar et al., 2008). Photoreceptors such as phytochromes have been shown to modulate responses to both biotic and abiotic stress (Carvalho et al., 2011; Indorf et al., 2007). Up-regulated transcripts mapping to *ABA-INDUCIBLE bHLH-TYPE (AIB)* and *MYC2* encode proteins that involved in regulating ABA-induced gene expression in Arabidopsis (Abe et al., 2003; Li et al., 2007), again indicating the influence of ABA signalling following salt shock in GD33DH. Transcripts mapping to the *LOTUS JAPONICUS Roothairless1-Like (LRL1 and LRL2)* TFs have been shown to positively regulate development of the tips of root hair cells and are controlled by auxin signalling (Tam et al., 2015). These genes were both up-regulated following salt shock treatment in GD33DH suggesting that the presence of increased salt in the soil affects the root architecture.

bZIP transcription factor family

The bZIP TF family is one of the largest TF families in plants and takes part in multiple processes, particularly abiotic stress responses, mediated through the ABA signalling pathway. The *cis*-acting ABRE element (ABA Responsive Element) and ABF TFs that bind the ABRE element (ABRE-binding protein/ABRE-binding factors) are bZIP which are key in ABA-dependent gene expression (Yoshida et al., 2014b). In this study there were 42 differentially expressed transcripts mapping to bZIP TFs, of which the ABA responsive genes *ABF3*, *ABF4*, *AREB3* and *ABI5* were up-regulated. ABF3 (Fig. 4.13e) and ABF4 proteins are part of trio of master regulators of ABA-induced gene expression in response to abiotic stress conditions (Yoshida et al., 2010). ABI5 is active during seed maturation and germination, regulating late embryogenesis-abundant genes during both developmental stages (Bensmihen et al., 2002). The observation that a selection of master regulators of the ABA dependent gene expression pathway were up-regulated under salt shock again highlights the importance of this signalling hormone in stress tolerance.

MYB and MYB-related transcription factor family

The functionally diverse MYB and MYB-related TF families constitute a large proportion of the differentially expressed TFs with 77 transcripts mapped to this family found to be differentially expressed. These TF families have been well characterised to have a role in developing tolerance to abiotic stresses (Li et al., 2015a), particularly drought stress (Baldoni et al., 2015). As described in Section 4.4.8, MYB96 (Fig. 4.13d) is ABA responsive and is involved in lateral root growth and in decreasing stomatal aperture in response to desiccation. Downstream targets include salt tolerance genes such as *RD22* in the leaf and the genes involved in the auxin pathway in the root (Seo et al., 2009). *MYB30*, *MYB60* and *MYB108* have been shown to be differentially expressed in response

to drought stress in *B. napus* (Liu et al., 2015a). MYB108 has a key role in the response to infection by *B. cinera* as well as roles in abiotic stress response and is induced in response to ABA, JA and ethylene (Mengiste et al., 2003). Here, the transcript mapping to this gene is down-regulated following salt shock, suggesting that MYB108 could be acting in its role of negative regulator of ABA induced death, as has been seen in response to wounding and in defence against necrotrophic pathogens (Cui et al., 2013). MYB60 is a regulator of stomatal movement and root growth and is down-regulated under drought stress (Oh et al., 2011). It is also down-regulated in response to salt shock in GD33DH, possibly resulting in an effect on root architecture and stomatal closure to protect the plant from further desiccation.

NAC transcription factor family

Plant specific NAC (NAM, ATAF1/2, and CUC2) TFs contain a highly conserved NAC DNA binding domain with variable C-terminal domains and are plant specific. They play key roles in plant development, senescence and abiotic and biotic stress responses (Breeze et al., 2011; Hickman et al., 2013; Windram et al., 2012; Xu et al., 2013b). In this experiment, 39 transcripts mapping to NAC TFs were identified as differentially expressed, indicating the importance of the TF family in the response to salt shock in GD33DH. Up-regulated in this experiment was the *NAC102* gene, which has a role in senescence (Breeze et al., 2011) and is activated by EIN2 (ETHYLENE INSENSITIVE 2) (Kim et al., 2014) and ATAF2, whose role was described in Section 4.4.8 (Fig. 4.13c). A transcript mapping to *NAC096* was down-regulated in response to salt shock. This is a surprising result given that a major proportion of abscisic acid (ABA) responsive genes are under the transcriptional regulation of NAC096, in response to dehydration and osmotic stresses in Arabidopsis (Xu et al., 2013b). The NAC096 protein has been shown to work together with ABF2 (a bZIP TF, which up-regulated in this study), activates important dehydration response genes such as *RD29* (Xu et al., 2013b). This could either be as a result of cross talk between the stress response pathways or this gene is present in triplicate in *B. oleracea* and the orthologs were not annotated, or not included on the array due to high levels of sequence similarity.

WRKY transcription factor family

Members of the WRKY protein family contain a highly conserved amino acid sequence motif WRKYGQK, responsible for the WRKY name. WRKY proteins act as transcriptional activators of key ABA-responsive genes e.g. ABI4, ABI5, ABF4, MYB2, DREB1A, DREB2A and RD29A. Members of the WRKY family also play an important role in defence and cross talk between signalling pathways (Deng-Hui et al., 2008; Li et al., 2010;

Wu et al., 2011). In this experiment, 19 transcripts mapping to WRKY genes were found to be differentially expressed in GD33DH in response to salt shock. Of these, the most well characterised is WRKY33, discussed in Section 4.4.8 (Fig. 4.13g). Other salt induced WRKYs include *WRKY15*, which is induced by oxidative and salt stress in Arabidopsis and the protein negatively regulates salt and osmotic stress tolerance (Vanderauwera et al., 2012). Also *WRKY28*, which is rapidly induced by ROS and the protein protects against fungal pathogens such as *B. cinera* that are known ROS producers (Wu et al., 2011).

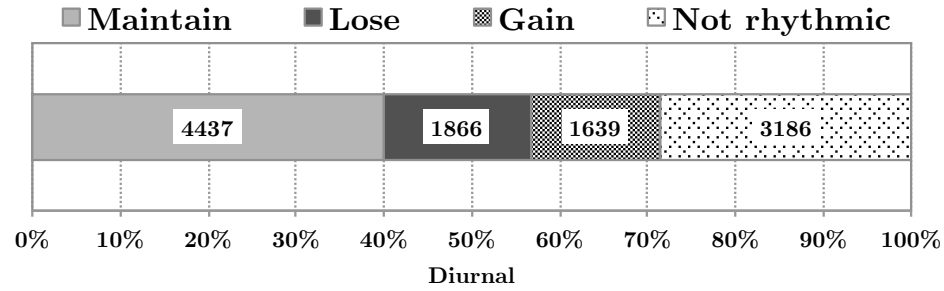
4.4.10 The effect of salt shock on circadian regulated transcripts

The response to abiotic stress has been directly linked to the circadian clock (reviewed in Grundy et al., 2015), and genes which are involved in the response to abiotic stress often show diurnal expression e.g. Figure 4.10 a and b. To investigate the effect that salt shock has on diurnal expression of the transcripts, the expression profiles of differentially expressed transcripts (both control and salt-treated) were analysed for diurnal expression patterns using the JTK CYCLE software (Hughes et al., 2010), as summarized in Figure 4.15a. By comparing treatment profiles to control, it was found that following salt shock, 39.8% of the transcripts maintained their diurnal expression pattern, as indicated in Figure 4.15b, which maps to Bo00975s030.1 an Inositol-3-phosphate synthase, which has been shown to enhance salt tolerance in a variety of crops (Abreu and Aragao, 2007; Sheveleva et al., 2002; Zhai et al., 2015). 16.7% of transcripts lost diurnal expression following salt shock as in Figure 4.15c which shows the expression profile of a transcript mapping to Bo9g098940.1, *ERD1* which follows the diurnal pattern of the control until 18 hpt, when expression increases in the treated samples. A small proportion of transcripts (14.7%) acquired a diurnal expression pattern, an example is shown in Figure 4.15d. This transcript maps to Bo1g098570.1 a chaperone DnaJ-like protein. The expression profile shows that the expression of the chaperone decreases in the salt-treated manner, then increases following a wave pattern, whilst the control does not. Finally, 28.8% of transcripts were not diurnally expressed, as demonstrated by Figure 4.15e which shows the expression profile of a transcript mapping to Bo8g042060.1, a transcription elongation factor protein which decreases in gene expression at 18 hpt in response to salt shock.

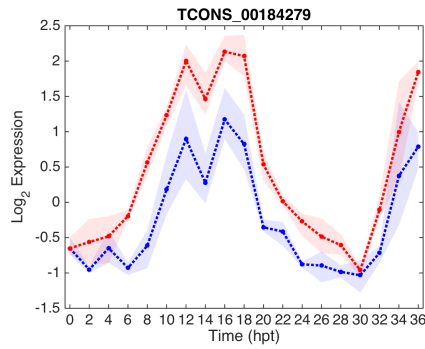
A diurnal expression pattern was altered (gained or lost) in 31.4% of transcripts (Figure 4.15a), suggesting that the response to salt shock is highly influenced by the circadian clock and by time of day effects.

4.4.11 Clustering differentially expressed transcripts

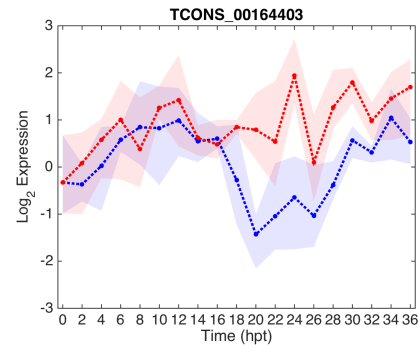
Analysing individual transcript profiles provides a highly detailed picture of the salt shock response in *B. oleracea* GD33DH. However, the sheer quantity of differentially expressed



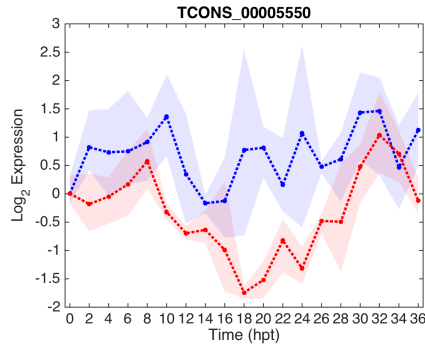
(a) Categories of diurnal expression present in differentially expressed transcripts



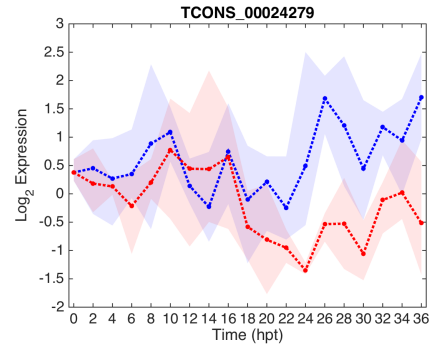
(b) Maintained rhythmicity



(c) Lost rhythmicity



(d) Gained rhythmicity



(e) No rhythmicity

Figure 4.15: Diurnally expressed transcripts following salt shock

Diurnally expressed transcripts following salt shock, as identified using the JTK CYCLE software. Patterns of diurnal expression amongst the differentially expressed transcripts is summarised in (a), in which the number in each category represents the number of transcripts belonging to each group. Transcripts were split into four categories (b) those which maintained diurnal expression; (c) those that lost diurnal expression following treatment with salt shock; (d) those that gained diurnal expression; (e) transcripts which were not diurnal expressed regardless of treatment. For the expression profiles (b - e), red corresponds to the salt-treated expression profile, blue to the control and shaded areas indicate minimum/maximum values.

Cluster size	Number of clusters
1-25	16
26-50	11
51-75	13
76-100	22
101-125	23
126-150	25
151-175	11
176-200	10
201-225	3
226-250	3
251-275	1

Table 4.4: Cluster size summary following MDI clustering

transcripts makes the inference of gene regulatory networks (GRNs) challenging. Thus, another method of analysing gene expression data was used, in which transcripts with similar expression profiles during salt shock were grouped together in clusters. This can be used to identify general trends in the data and to identify groups of potentially co-regulated transcripts that are important in the response to salt shock. The 11,754 differentially expressed transcripts, using expression data for both salt treated and control were clustered together using the Multiple Dataset Integration (MDI) algorithm (Kirk et al., 2012; Mason et al., 2016; Savage et al., 2013). This method was used to produce 145 clusters, a subset of which can be seen in Figure 4.16. As indicated in Table 4.4, the size of the clusters was evenly spread, with the majority containing between 75 and 200 transcripts. Zero-centred, standard deviation normalised expression profiles of clusters are given in Appendix F and a selection of clusters are shown in Figure 4.16.

Using the closest Bo gene ID for each transcript, the clusters were analysed for over-representation of GO terms using BiNGO (Ashburner and Lewis, 2002; Maere et al., 2005b) to determine whether the closest Bo gene ID mapping to the transcripts present within the same cluster were involved in related biological processes, suggesting co-ordinated regulation. Clusters were associated with a diverse range of GO terms, with 95/145 clusters having multiple enriched GO terms. The most significantly over-represented GO terms for each cluster are given in the Appendix G, and a selection are shown alongside a selection of clusters with an interesting shape in Figure 4.16. The top over-represented GO terms for the clusters were associated with metabolite processes such as photosynthesis and cellular biosynthetic processes, growth processes such as epidermal cell differentiation, transport and hormone signalling (ABA mediated signalling, JA and ethylene-dependent systemic resistance).

Looking at all of the cluster shapes (Appendix F), it is again clear that the circadian clock plays a key role in the regulation of transcripts responding to salt shock, with at least

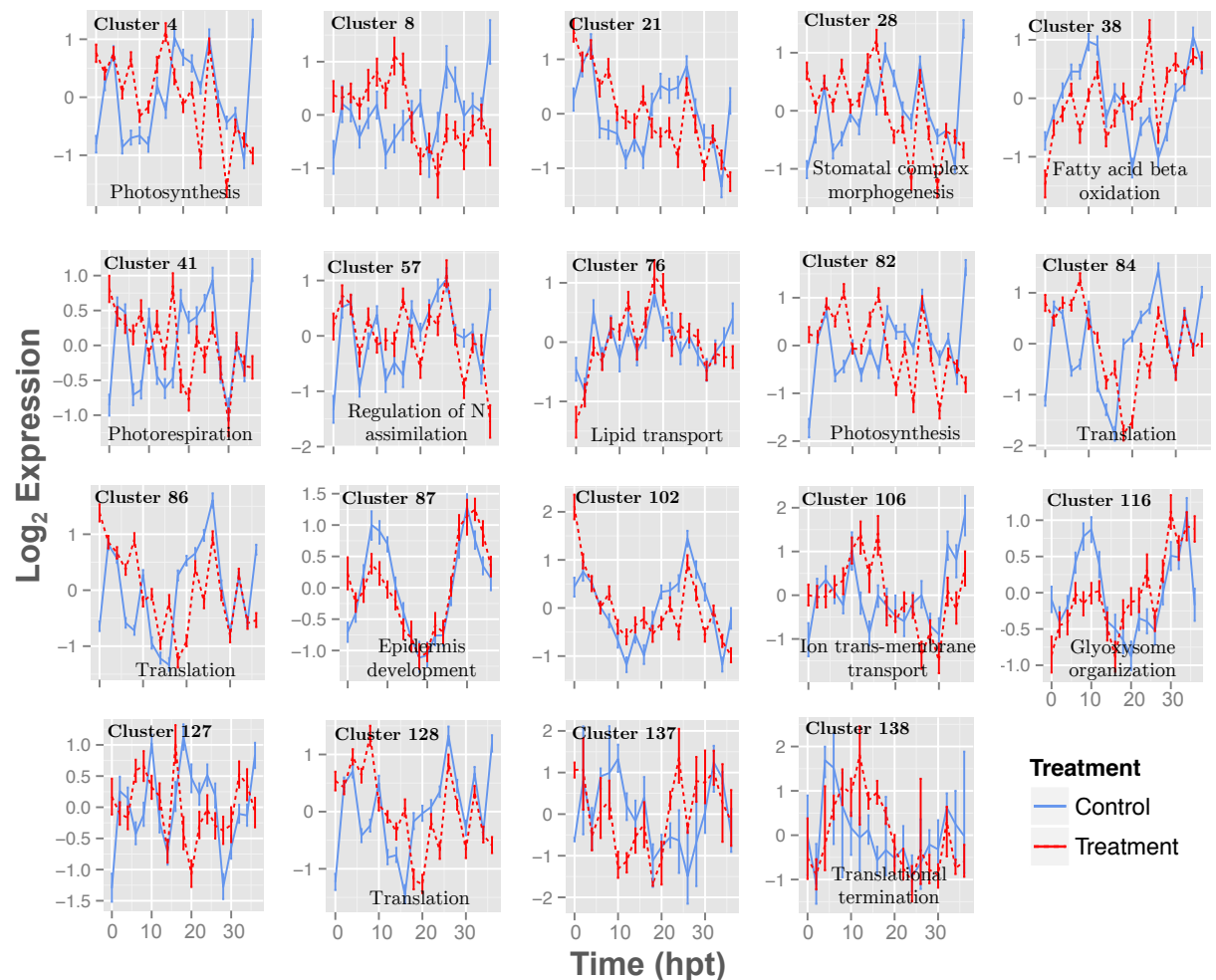


Figure 4.16: Plots of the mean gene expression profile of a selection of clusters of co-expressed transcripts differentially expressed after salt shock in *B. oleracea* GD33DH

Both treatment and control expression data was clustered alongside using the Multiple Data Integration algorithm. The red line represents the mean expression profile for the salt-treated transcripts and blue the control. Data were mean centred and standard deviation normalised separately for each condition prior to clustering. The error bar is the 99% confidence interval of the data within each cluster. Note scale may be different for each plot. The x -axis is hours post treatment (hpt) and the y -axis is the Log_2 Expression. Enriched GO terms, if present are given at the bottom of the cluster plot.

55% of cluster profiles showing clear diurnal expression. That control and salt-treated expression profiles were clustered together means that expression profiles can be further separated by the differences between the treatment and control profiles caused by the stress condition. For instance in the example clusters in Figure 4.16, Clusters 4, 21, 38, 86, 87, 102, 106, 116, 128 and 137 show where diurnal expression is lost or disrupted following salt shock, compared to the control expression profile. These clusters are enriched with a diverse range of GO terms, such as ‘photosynthesis’, ‘fatty acid beta oxidation’, ‘translation’ and ‘epidermis development’ indicating that a vast range of biological processes under the influence of the circadian clock are affected by salt shock.

Many metabolic processes are disrupted in salt-treated plants compared to control as indicated by the broad nature of GO terms enriched in the clusters. Indeed, ‘translation’ is the most prevalent GO term, present as the top GO term in 8 clusters (see Appendix G) which indicates that salt shock disrupts the cells ability to produce new proteins as ribosomes are highly sensitive to a high salt environment (Omidbakhshfard et al., 2012), as shown in Clusters 15, 79, 84, 86, 96 and 138 (Fig. 4.16). Photosynthesis is also greatly affected by salt shock, and is the top GO term for Clusters 4 and 82 (Fig. 4.16) which show down-regulation in the salt treated samples compared to control. Other clusters are also associated with photosynthesis through highly related GO terms such as ‘regulation of photosynthesis’, ‘protein-chromophore linkage’, ‘chlorophyll metabolic process’ and ‘generation of precursor metabolites and energy’ being prevalent in the clusters (see Appendix G) indicating the disruption caused to photosynthesis by salt shock.

Damage to the lipid membrane of the cell and transport of sodium ions is evidently taking place as many clusters present are associated with fatty acid beta-oxidation, organisation of the lipid membrane and transport of ions, as shown in example Clusters 38, 76, 106 and 116 (4.16) and other clusters as seen in Appendix G. Cluster 116 shows up-regulation of genes associated with ‘Glyoxysome organisation’ a type of peroxisome containing enzymes that are involved in lipid mobilisation via beta oxidation of fatty acids. They are usually present in post germinative seedlings, however their presence has been observed in senescent leaves (Donaldson et al., 2001). GO terms relating to ‘cell death’ and ‘autophagy’ are seen in the clusters (Appendix G), suggest that the plant maybe unable to cope with the high level of sodium ions and initiates cell death in older, fully expanded leaves (fully expanded leaf #5 was sampled in this experiment) as growth of new tissue is not occurring in these leaves.

In the example clusters, there are several clusters without enriched GO terms that have interesting expression patterns, This includes clusters that experience a clear decrease in transcript expression either at a given time point (eg Cluster 8, 18 hpt and Cluster 84, 8 hpt) or gradually throughout the time-course (Clusters 4, 21 and 28). Some clusters also show up-regulation of gene expression, for instance Cluster 76 sees up-regulation of stress

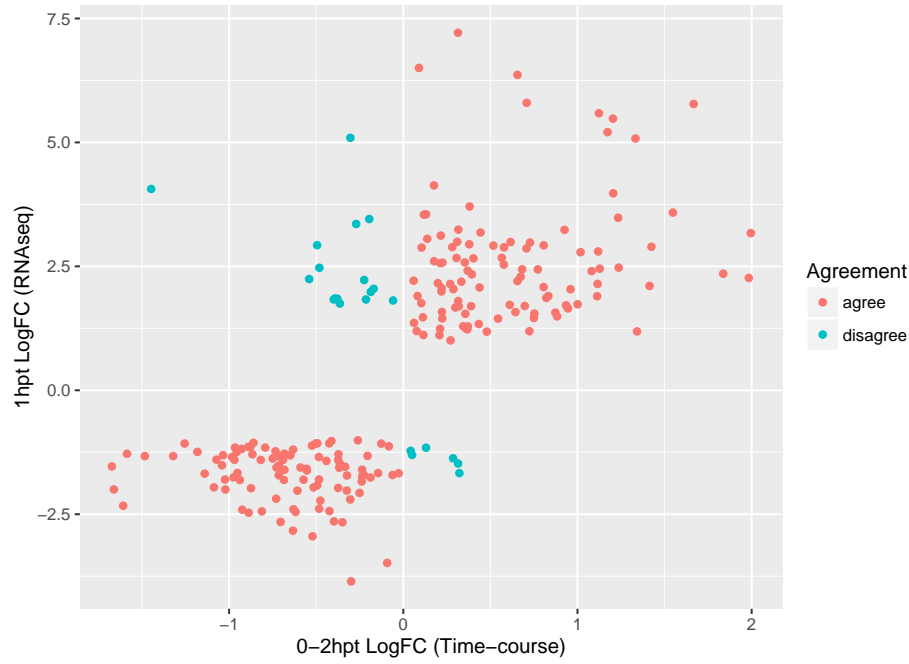


Figure 4.17: Correlation between difference between treatment and control in time series experiment and logFC of gene expression in salt shock RNAseq data

Red indicates instances where the direction of expression agreed and blue indicates instances in which the direction of expression did not agree between the two datasets. The x -axis is the logFC of genes differentially expressed between 0 and 2 hpt in the time-series dataset. The y -axis is the logFC for the RNAseq experiment sampled at 1 hpt.

related transcripts, this cluster contains two of the top 20 most differentially expressed transcripts, though the effect of the difference in expression seen in the profiles (Fig. 4.6) is lost in the clustering due to the normalisation methods used to produce the cluster plots (zero-centred, standard deviation normalised). Cluster 116 shows a gradual increase throughout the time-course, as does Cluster 38.

The cluster analysis of differentially expressed transcripts is highly informative. Images comparing treated and control samples show a wide range of different patterns and clustering allows transcripts that may be co-regulated and involved in highly co-ordinated processes to be identified.

4.4.12 Validation of differentially expressed transcripts using RNAseq data

To validate the time series data, the differentially expressed transcripts were compared to the differentially expressed genes in the RNAseq experiment described in Chapter 3. In this RNAseq experiment plants exposed to salt shock (500mM) and a control were harvested at 1 hpt ($n=3$). The methodology of this experiment differed from the time

series experiment in that the treatment was a higher concentration of saline (500mM) and sampling occurred at 1 hpt, where as in the time series experiment the earliest sampling point was 2 hpt. All other factors such as growth conditions and method of application remained the same. In order to validate the differentially expressed transcripts, the transcript data was summarised to the Bo gene ID level and compared with the output of the RNAseq experiment, in which RNAseq reads were aligned to the TO1000 genome and hence have Bo gene IDs also.

Analysis of the RNAseq data resulted in 1,777 differentially expressed genes of which 1,082 genes were up-regulated and 695 genes were down-regulated, as determined using the direction of the logFC. Out of 2,805 genes that were differentially expressed at 0 or 2 hpt in the time series experiment (identified using the Time Of Differential Expression analysis, GP2S time local method there were 222 overlapping differentially expressed genes in the RNAseq validation experiment (Stegle et al., 2010; Windram et al., 2012; Fig. 4.8a). The result was a relatively low number of differentially expressed genes, however sampling occurred early, before the first time point of the time series experiment meaning it may have been too early to detect a large proportion of differentially expressed genes.

GeneID	Description	LogFC (Validation)	LogFC (2hpt Time-course)	Average LogFC (Time-course)	TOFDE
Top 20					
Bo9g059510.1	Acid phosphatase 1, putative	-2.83	-1.36	-1.58	0h
Bo2g148140.1	Ninja-family protein	5.77	2.04	1.42	2h
Bo6g064670.1	NAC domain containing protein	4.06	-2.51	-1.48	2h
Bo7g087520.1	Ninja-family protein	7.21	0.40	2.36	2h
Bo9g031100.1	Cystathionine gamma-lyase	2.35	3.05	2.18	2h
Bo8g049770.1	Abcisic acid receptor PYR1	-1.95	-1.291	-1.41	2h
Bo9g022010.1	GATA transcription factor	-1.40	-0.86	-0.56	2h
Bo5g022200.1	Conserved hypothetical protein	2.80	2.40	1.28	2h
Bo3g023690.1	Non-specific lipid-transfer protein	NS	0.49	4.04	0h
Bo1g047810.1	NADH:ubiquinone oxidoreductase subunit	NS	2.52	2.08	0h
Bo5g013650.1	S-receptor kinase	NS	-1.30	-0.85	0h
Bo3g165670.1	Senescence-inducible stay-green protein	NS	0.85	1.21	2h
Bo9g027620.1	Threonine synthase, putative	NS	-1.48	-1.31	2h
ABA biosynthesis					
Bo3g066190.1	NCED3	5.47	0.93	0.79	2h
Bo9g020440.1	Zeaxanthin epoxidase	1.11	0.17	0.50	2h
Bo6g028000.1	Xanthoxin dehydrogenase	-2.24	0.32	0.21	4h
Bo4g161370.1	Aldehyde oxidase	NS	0.47	0.26	14h
Key genes					
Bo9g014980.1	Myb domain protein	3.48	2.25	0.91	2h
Bo1g007700.1	ABA responsive elements-binding factor	2.92	0.97	0.69	2h
Bo5g030290.1	Dehydrin	NS	0.001	1.21	2h
Bo3g142840.1	Zinc-finger protein	NS	2.34	0.93	8h
Bo9g098940.1	Chaperone clpB 1	NS	1.44	0.55	14h
Bo4g190900.1	Dehydration-induced protein (ERD15)	NS	0.65	0.43	16h
Bo9g011300.1	BURP domain-containing protein	NS	0.84	0.29	16h
Bo5g086990.1	bHLH family transcription factor	NS	0.46	0.27	18h
Bo3g032500.1	WRKY transcription factor	NS	1.04	0.42	20h

Table 4.5: Validation of selected genes in time series experiment and RNAseq experiment

A comparison of a selection of ABA biosynthesis genes, key marker genes and top 20 differentially expressed transcripts, that have been discussed in this chapter, to an RNAseq salt shock experiment sampled 1 hpt. LogFC is given for the RNAseq validation experiment, at 2 hpt in the time-course experiment and the average logFC of all time points in the time series experiment. TOFDE stands for Time Of First Differential Expression, as determined using the GP2S-TL output.

Unlike the time series experiment which showed a majority of down-regulation, there were more up-regulated genes (120 genes) compared to down-regulated genes (102 genes). This could be due to the early sampling time or due to the inherent differences in using RNAseq to measuring gene expression compared to microarrays, as RNAseq does not rely on prior knowledge to design probes thus the outcome in theory, is less constrained.

In order to validate the patterns of expression seen in the time series experiment, the direction of change (i.e. up/down-regulation) of the differentially expressed genes at 0-2hpt were checked for agreement between the two experiments (Fig. 4.17). Looking at the overlapping differentially expressed genes, when the logFC from the RNAseq experiment was compared with the logFC at 0 - 2 hpt in the time-course experiment, there were 200 genes in which the direction of expression was in agreement (Fig. 4.17, red) and 22 genes where the direction of expression was not in agreement (Fig. 4.17, blue). There was a relatively high agreement between the two datasets given, that expression was measured using different technologies, with a Spearman correlation of 0.58 which sufficiently supports the integrity of the data.

Of the top 20 transcripts in the time series (Fig 4.6) mapping to Bo gene IDs, 8 Bo genes were differentially expressed in both experiments, and showed good agreement with regards to the direction of expression, except in the case of Bo6g064670.1 (NAC domain containing protein) which was down-regulated in the time series experiment and up-regulated in the RNAseq experiment.

When the ABA biosynthesis genes were examined, Bo3g066190.1 (encoding 9-cis-epoxycarotenoid dioxygenase, *NCED3*) and Bo9g020440.1 (Zeaxanthin epoxidase, *ABA1*) were up-regulated in the RNAseq experiment, whilst Bo4g161370.1 (aldehyde oxidase, *AAO3*), which is important in the ABA biosynthesis pathway, was not up-regulated in the validation experiment (Table 4.5). The results can be made more clear by analysing the expression profiles of the ABA biosynthesis genes (Fig. 4.10) and by looking at the time of first differential expression (Table 4.5). It can be shown that *NCED3* and *ABA1* are first differentially expressed 2 hpt and *AAO3* was up-regulated at 14 hpt, indicating that it was simply too early for this gene to be up-regulated in the validation experiment. Interestingly, Bo6g028000.1 (Xanthoxin dehydrogenase, *ABA2*) was down-regulated in the validation experiment, but found to be up-regulated at 4 hpt in the time-course experiment.

Of the marker genes, *MYB96* and *ABF3* (encoded for by Bo9g014980.1 and Bo1g007700.1, respectively; Fig. 4.13) are differentially expressed in both experiments. Other key genes were not found to be differentially expressed in the validation experiment. Given the time of first differential expression (Table 4.5) this is expected as many of the genes listed here are not up-regulated in the validation experiment, are up-regulated much later in the time-course, between 8 and 20 hpt.

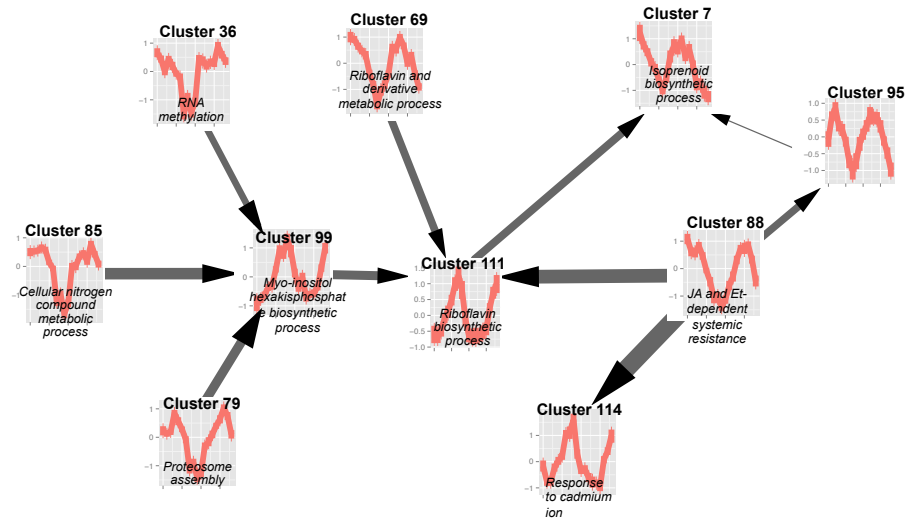
The output of the validation experiment strongly support the results of the high-

resolution transcriptomic time series experiment, as a number of genes were differentially expressed in both experiments, the majority in the same direction suggesting similar functions (Fig. 4.17). Genes that were differentially expressed in the time-course at later points were not found to be differentially expressed in the validation experiment (Table 4.5) giving confidence to the output of the time of differential expression analysis (Fig. 4.8).

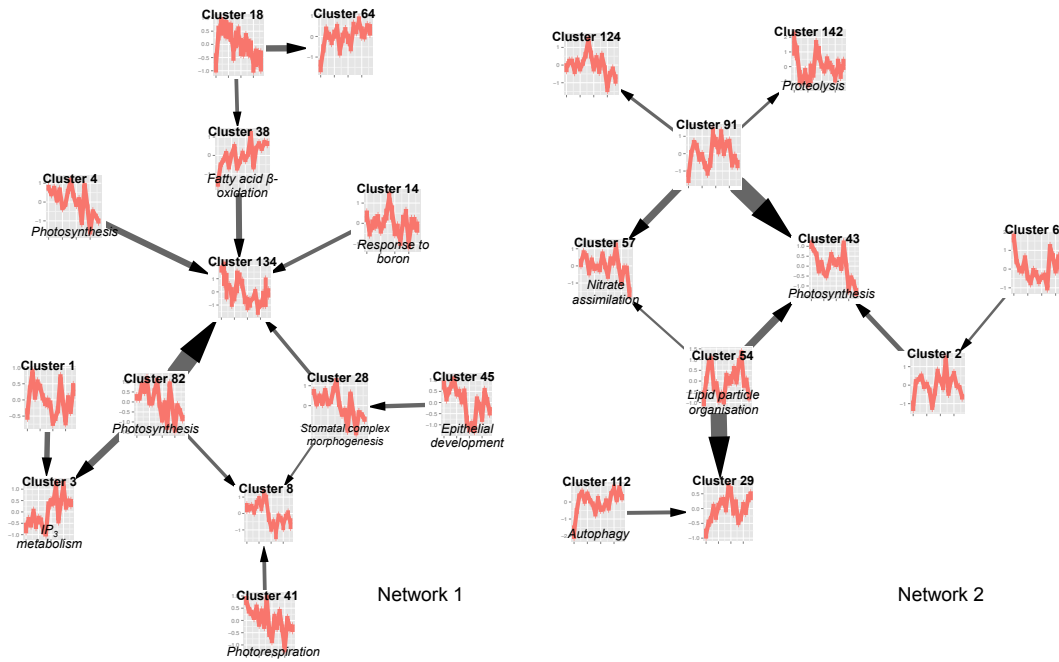
4.4.13 Inferring regulatory networks

To identify how co-expressed transcripts interact and are regulated, regulatory networks can be inferred from the time series expression data using the Causal Structure Inference (CSI) algorithm (Penfold et al., 2012, 2015). The number of individual differentially expressed transcripts is far too large to produce a transcriptome-wide network model, therefore the mean of the treatment data for each cluster was used as a representative group of transcripts for network inference. The network model produced by modelling the salt-treated cluster means using CSI is shown in Figure 4.18a and b, alongside significantly over-represented GO terms for clusters with the strongest marginal probabilities. Cluster 88 was the main regulatory cluster, regulating Clusters 95, 111 and 114. Interestingly, cluster 88 did not contain any TFs, however it contained many highly differentially expressed protein kinases suggesting a key role for these kinases in the transduction of the ABA stress signal. Cluster 95, regulated by cluster 88 contains several important TFs, including a bHLH with orthology to *ICE1* and transcripts with homology to *HB6* and *WRKY28*. WRKY28 has been shown alongside bHLH17 to confer resistance to abiotic stress conditions such as oxidative stress (Babitha et al., 2012) and also in resistance to infection with *B. cinera* (Wu et al., 2011), both in Arabidopsis. Cluster 111, containing 64 transcripts, was regulated by Clusters 69, 88 and 99 and regulated Cluster 7. Cluster 111 contained several transcription factors, including two MYB TFs, an AP2-like ethylene transcription factor with orthology to TARGET OF EAT 3 (TOE3) and an ethylene-responsive transcription factor which is a floral repressor regulated by miRNA172 (Jung et al., 2014) suggesting a role in the repression of development of GD33DH under salt shock.

From Figure 4.18a it was clear that diurnal expression is the key feature of this model, suggesting that perhaps clusters with strong diurnal patterns may be eclipsing key clusters without diurnal expression, leading to an incomplete model in which the regulatory roles of genes within clusters cannot be established. Therefore, the diurnal nature of each cluster was determined using JTK Cycle and non-diurnally regulated clusters were selected for modelling, as shown in Figure 4.18b. Modelling using non-diurnal clusters resulted in the formation of two regulatory networks suggesting multiple regulatory networks are in place under salt shock. In network 1 (Fig. 4.18b, left) Clusters 28 and 82 appeared to be regulating a number of clusters. Interestingly, these clusters contained



(a) Network inference of cluster means



(b) Network inference of non-diurnal cluster means

Figure 4.18: Inferred network models using the Causal Structure Identification algorithm

(a) The mean expression profile for all clusters and (b) non-diurnal clusters was taken and modelled using the Causal Structure Identification (CSI) algorithm (Penfold et al., 2012, 2015). Each node represents a cluster produced from the Multiple Data Integration algorithm, clustering of transcripts differentially expressed in response to salt shock in GD33DH. For each cluster, the red line represents the mean expression profile for the salt-treated transcripts. Note scale may be different for each plot. The top GO term is given in italics below the expression profile. The arrows indicate the direction of regulation and the thickness of the grey line indicates the marginal probability.

TFs involved in the regulation of cytokinin signalling. Cluster 28 contained a transcript which mapped to Bo7g109100.1 (*CYTOKININ RESPONSE FACTOR 2*, *CRF2*) that was highly down-regulated. Cluster 82 contained transcripts mapping to a down-regulated response regulator (Bo9g045370.1). Response regulators and CRFs have roles in the cytokinin signalling (Ishida et al., 2008; Nguyen et al., 2016b; Rashotte and Goertzen, 2015) suggesting the importance of cytokinin signalling in regulating the response to salt shock in GD33DH (Nishiyama et al., 2012). In network 2 (Fig. 4.18b, right), Clusters 54 and 91 were shown to regulate a number of other clusters involved in key processes such as proteolysis and photosynthesis. Cluster 54 contained an up-regulated HB TF mapping to Bo7g096160.1 (*KNAT3*), which is repressed by moderate levels of cytokinin (Truernit et al., 2006). Cluster 91 is a small cluster of 43 transcripts and contained transcripts mapping to Bo6g119860.1, an abscisic acid responsive elements-binding factor. AREBs are heavily involved in regulation of ABA responsive genes (Yoshida et al., 2010) suggesting cross talk and cross regulation between various signalling pathways.

The results of the CSI network inference highlight the usefulness of these data in network inference and identification of potential key regulatory and downstream genes involved in response to salt shock in GD33DH.

4.5 Discussion

Many transcripts associated with multiple biological pathways and functions were identified in the time series analysis, revealing a complex transcriptional network controlling the response to salt shock in GD33DH.

A global transcriptional reprogramming takes place in GD33DH responding to salt shock. Between 0 - 36 hpt a total of 11,754 transcripts were identified as differentially expressed. These transcripts mapped to 7,141 Bo gene IDs, representing a significant change in 12% of the GD33DH transcriptome in response to salt shock. In addition to the transcripts mapping to a Bo gene ID, there were 1,573 differentially expressed transcripts that were differentially expressed that did not map to any Bo gene ID suggesting potential novel transcripts relating to salt tolerance in GD33DH. Several of these transcripts showed a high level of differential expression, as is indicated by the inclusion of 2 in the top 20 differentially expressed transcripts (Fig. 4.6).

Two methods of identification were used to determine differential expression, standard *F*-tests for equality of variance, constructed from the fitted models for each transcript and GP2S, which ranks transcripts according to the degree of differential expression and is also capable of determining the time at which transcripts first become differentially expressed. Both methods have advantages and disadvantages and as such, the intercept of each method was used to determine a robust list of differentially expressed transcripts in

response to salt shock.

The F -test partitions variation into terms specified by the model. In this experiment, variation was partitioned between time and treatment to determine differential expression. F -tests comparing variance in the treatment term resulted in 94,232 (over 50% of the total transcripts present on the array) differentially expressed transcripts, which is an extremely large number of differentially expressed transcripts to work with. Standard F -tests are not designed for use with time series data as they assume independence, however given the nature of a time series experiment there is likely to be no independence between time points. F -tests are also sensitive to deviations from normality. Given that differential expression is based on variance, noise in the data between technical and biological replicates although reduced following model fitting, is likely to have a substantial impact on differential expression resulting in a potentially large number of false positives in the data.

GP2S determines differential expression by fitting either one or two Gaussian processes to the data and determine which provides the best fit via a Bayes factor score (likelihood of differential expression), as previously described. In this experiment, a Bayes factor of ≥ 14 was used, which is extremely conservative compared to other studies e.g. a Bayes factor score of ≥ 6 was used in a progressive drought study in Arabidopsis and the general advice for using Bayes factors is that a Bayes factor ≥ 10 is considered to be strong evidence for differential expression (Bechtold et al., 2016; Calderhead and Girolami, 2009; Kass and Raftery, 1995), therefore using a Bayes factor of ≥ 14 resulted in a conservative list of 13,638 differentially expressed transcripts. Given the vast difference in number of differentially expressed transcripts between both methods, the intersection was taken, resulting in a list of 7,141 differentially expressed transcripts in GD33DH in response to salt shock. It is likely that using the intersection of both methods has resulted in the removal of a number of true positives from the final list of differentially expressed transcripts, however using multiple methods to confirm differential expression gives confidence in the output of the transcriptomic analysis, strengthening biological conclusions.

Published analyses of both biotic and abiotic stress treatments in different plant species have resulted in a similar proportion of transcriptional reprogramming. In *B. napus* responding to drought stress, a total of 6,018 and 5,377 differentially expressed genes were identified in root and leaf, respectively (Liu et al., 2015a). In Arabidopsis, infection with *B. cinera* identified 9,838 differentially expressed genes (Windram et al., 2012), the senescence process in Arabidopsis has been shown to involve 6,323 differentially expressed genes (Breeze et al., 2011) and finally 5,545 genes were differentially expressed in response to salt stress in a salinity-tolerant genotypes of chickpea (Garg et al., 2016).

4.5.1 Clustering based on time of first differential expression

In order to make meaning biological conclusions from the differentially expressed transcripts, two methods of clustering were used to group the data based on time at which the transcript was first differentially expressed and based on the shape of the transcript expression profile. The first method of clustering allowed for the establishment of a chronology of transcript expression in the salt shock response. Using GO term analysis it was then possible to link biological function to transcript expression. This also allows for analysis of individual groups of transcripts involved in a diverse range of biological functions such as ion homeostasis, transcription factor families, hormone related gene expression, photosynthesis and metabolism. It was clear from both GO term analysis (Fig. 4.9) and by investigating the number of differentially expressed transcripts at each time point (Fig. 4.8) that the osmotic and ionic phases described by Munns and Tester (2008) were present in GD33DH in response to salt shock, albeit at a much more rapid pace than is experienced in salt stress, presumably due to the severity of the shock (Shavrukov, 2013).

Ion homeostasis

The regulation of transport of ions across the plasma membrane is well characterised in response to abiotic stress conditions. In the time series experiment, differentially expressed ion transporter related transcripts included ABC transporters, aquaporins, MATE efflux proteins, CNGCs, V-type proton ATPases, ABC transporters and transcripts with orthology to NHX1 which has roles in maintaining Na⁺ and K⁺ homeostasis. Interestingly, the majority of these transporters were down-regulated, including 3 sodium/hydrogen exchanger proteins which are involved in the sequestration of Na⁺ and K⁺ ions in the vacuole. That transcripts belonging to the same family e.g. potassium transporters were found to have members both up- and down-regulated suggests that there is divergence in function within closely related transcripts.

The largest two groups of differentially expressed transporters, belonging to the Multiple Drug Resistance (MDR) gene family were the MATE efflux proteins and ABC transporter proteins (Table 4.3). Many genes within this family are yet to be characterised (Remy and Duque, 2014), however these proteins have been implicated in the salt stress response (Jiang and Deyholos, 2006; Li et al., 2015b; Sengupta et al., 2015) and it is possible that their role has previously been under-appreciated in the response to salt shock.

ABC transporters have roles in development and survival and can transport stress-related secondary metabolites and hormones such as alkaloids, terpenoids, polyphenols, quinines, ABA and auxin (as reviewed in Kang et al., 2011). The role of ABC transporters in response to salt stress has been investigated in rice (Sengupta et al., 2015) and an ABC transporter protein was found to be up-regulated in *G. hirsutum* under salt stress

conditions (Li et al., 2015b). Through knockout studies in Arabidopsis, it has been shown that an ABC transporter had roles in maintaining a high $K^+ : Na^+$ ratio through K^+ uptake in the salt stress response (Lee et al., 2004). The majority of ABC transporters are down-regulated in this experiment (8 up-regulated, 25 down-regulated), thus manipulating expression of these proteins may be ideal targets for enhancing salinity tolerance.

Multidrug and Toxic compound Extrusion (MATE) efflux proteins have roles in uptake and storage of specific compounds such as proanthocyanidin and epicatechin 39-O-glucoside (Zhao and Dixon, 2009), development and the response to stress (Tiwari et al., 2014). There is an equal number of up- and down-regulated MATE efflux protein in response to salt shock in GD33DH, suggesting diverse functions for this gene family.

Based on the transcripts mapping to transporters that were differentially expressed, GD33DH potentially uses three mechanisms for the maintenance of ion homeostasis following salt shock. That several transcripts were located on the vacuole such as V-type ATPases suggest that GD33DH attempts to store excess Na^+ ions within the vacuole to protect the cytoplasm from potentially harmful toxicity caused by excess ions. A large number of differentially expressed K^+ transporters suggests that the cell attempts to maintain a high $K^+ : Na^+$ ratio, which has previously been associated with increased tolerance to high salinity conditions. In addition, many transcripts mapping to transporters may have been down-regulated potentially preventing entry of excess Na^+ ions into the cytoplasm via an electrochemical gradient.

Hormone signalling and regulation by TFs

Transcripts implicated in hormone signalling, primarily ABA signalling, but also auxin, brassinosteroid, cytokinin, ethylene, GA, SA and JA were found to be differentially expressed in GD33DH responding to salt shock. Transcriptional regulation of the gene networks was potentially under the control of a range of TF families including AP2-EREBP, bHLH, bZIP, HB, MYB and NACs, many of which have been previously associated with the response to abiotic stress conditions. In addition to ABA biosynthesis related transcripts, also identified were differentially expressed TF families involved in hormone biosynthesis and response, including auxin (IAA and ARF), ethylene (AP2-EREBP), gibberellin (GRAS) and cytokinin (ARR). The role of hormone signalling, particularly ABA in response to salt stress conditions has been extensively demonstrated in the literature (Peleg and Blumwald, 2011; Raghavendra et al., 2010; Verslues, 2016; Yoshida et al., 2014b). Transcripts mapping to TF families involved in the regulation of stress response (*bZIP*, *bHLH*, *HB*, *MYB*, *NAC* and *WRKY*) as well as developmental processes (*ARF*, *MADS*). These results indicate crosstalk between multiple pathways resulting in complex transcriptional networks controlling downstream cellular responses to salt shock (Chan, 2012; Hartmann et al., 2015; Seki et al., 2002). The results of the network inference of the ABA signalling pathway

suggest a handful of genes that may be suitable candidates for manipulation such as *PYR* genes Bo3g022900.1, Bo4g027570.1, Bo5g115300.1 and Bo7g075740.1. These are highly connected, sit towards the top of the signalling network and are proposed to regulate multiple down-stream genes in the salt shock response. Over-expression of *PYR* ABA receptors has been shown to confer enhanced response to ABA and plant drought resistance in *Arabidopsis* (Gonzalez-Guzman et al., 2012) suggesting potential in the development of stress tolerant crop plants.

Growth, metabolism and photosynthesis

Based on chronology of the salt shock response, growth is rapidly down-regulated (at 0 hpt) followed by metabolism from 4 hpt and photosynthesis at 18 hpt.

Growth of GD33DH following salt shock conditions is reduced. Cell expansion of the shoot is rapidly inhibited via a calcium signal which is initiated at the cell membrane once the concentration of Na^+ rises above a certain threshold (Kader and Lindberg, 2010). This signal is propagated via the cell membranes from root to shoot (Kurusu et al., 2015). In *Arabidopsis*, this takes as little as two minutes (Choi et al., 2014). Evidence for this reduction in growth at the transcript level is through the over-representation of GO terms such as ‘multidimensional cell growth’, ‘tissue/cell development’ and ‘root and shoot morphogenesis’ in down-regulated transcripts. Later in the time series, down-regulation of transcripts relating to cytokinin stimulus, may also contribute to the reduced growth seen in GD33DH as a decrease in cytokinin has been linked with reduced growth (Nishiyama et al., 2012). This may indicate a potential area for manipulation in the development of salt-tolerant *Brassica*, as has been seen in several other species, both through exogenous application (Akter et al., 2014; Ghorbani Javid et al., 2011) and endogenous manipulation of the cytokinin biosynthesis pathway. However, thus far this has resulted in conflicting results (Kang et al., 2012; Le et al., 2012; Peleg et al., 2011; Wang et al., 2015).

Metabolism and protein biosynthesis are energy expensive processes. Based on expression of related transcripts, a trade off between metabolism must occur in order to promote tolerance mechanisms. It is likely that the plant shuts down these processes to redirect energies into protecting the cells from further damage caused by increase Na^+ . The vast number of down-regulated ribosome subunit proteins indicated that protein biosynthesis was also down-regulated under salt shock treatment. A decrease in the protein content of cells following salt stress is seen in *Arabidopsis* (Ndimba et al., 2005), and is consistent with the over-represented GO terms which are down-regulated at 10 hpt.

Transcripts relating to photosynthesis were down-regulated at 18 hpt. This may be caused by various factors. Firstly, changes in stomatal behaviour due to the build up of ABA within the guard cells, may affect gas exchange and the availability of CO_2 for photosynthesis (Stepien and Johnson, 2008). Secondly, due to changes in osmotic potential

within the system, water may also not be readily available to the cell for photosynthesis. Finally, increased Na^+ ions may lead to the breakdown of chlorophyll (Ashraf and Harris, 2013; Chaves et al., 2008), as indicated by the increased expression of a transcript mapping to the senescence-inducible chloroplast stay-green protein (*SGR1/NYE1*) (Sakuraba et al., 2014b), which was found in the top 20 differentially expressed transcripts.

Conclusions drawn from transcriptomic data must be made with caution, as the transcriptome changes may not necessarily translate to changes in protein levels due to the post-transcriptional regulation of mRNA and protein. In addition, biological conclusions drawn from GO term analysis must also be tentative as not all genes carry ontology annotations, as discussed in Chapter 3.

Clustering transcripts in this manner is extremely useful for analysis discrete groups of genes and for determining a chronology to the salt shock response. However, it is not possible to determine key regulatory genes and gene regulatory networks when the data is clustered based on time of differential expression, therefore another method of clustering was carried out, as below.

4.5.2 Clustering based on expression profiles

Another method of analysis of differentially expressed transcripts in high-resolution time series experiment is to cluster transcripts based on the shape of expression profiles (Bechtold et al., 2016; Breeze et al., 2011; Windram et al., 2012). This has advantages in being able to identify genes which are co-expressed as they will fit into the same cluster. Generally clustering is only carried out on the expression profiles of a singular condition i.e. the treatment expression profiles. It was interesting to take advantage of new clustering algorithms such as Multiple Dataset Integration (Mason et al., 2016) to cluster both control and treatment expression data so that interesting gains and losses between control and treatment expression profiles were identified. This was advantageous in that a large number of clusters were identified (145 clusters) containing co-expressed transcripts and using GO term analysis, biological function of each collection of genes could be established (Fig. 4.16). Whilst clustering based on time of first differential expression had advantages in uncovering discrete biological functions, clustering based on expression profile was more useful in terms of gaining a bigger picture of events, for instance determining the effects of the circadian clock on stress response genes, and also for inferring regulatory networks as discussed below.

Network inference using high-resolution time series data

An advantage of high-resolution transcriptomic time series analyses is the amenability of the data for network inference to identify regulatory transcripts. Modelling with cluster

means allowed for the potential identification of key regulatory transcripts that influence a range of other clusters, making them potential targets for crop improvement. However, in this experiment the diurnal expression pattern of the clusters was too strong to draw any meaningful biological conclusions from the modelling output. A large proportion ($\sim 50\%$) of the *Arabidopsis* genome is under circadian control under abiotic stress conditions (Covington et al., 2008). Differences in the time of day induction of drought stress in *Arabidopsis* suggests that plants are primed for certain stress conditions at different points in the day-night cycle, with different groups of genes induced by the same stress treatment given at different times throughout the day (Wilkins et al., 2010). When these diurnally expressed clusters were removed from the modelling, more informed gene regulatory networks could be established in which regulatory elements and regulation of cross talk between hormone signalling pathways could be seen. Another approach would be to remove the diurnal effect from the treatment data by normalising against the control, however this was beyond the scope of this thesis. It must be taken into consideration that the computational model has been constructed based on noisy expression data would need to be thoroughly validated through the use of knockout mutants before any clear biological conclusions could be drawn.

The clustering methods presented here are complementary to each other, and each provide a unique perspective in divulging transcriptomic changes, through analysis of individual groups of transcripts in order of expression and to establishing regulatory transcripts in response to salt shock. Using both methods in the analysis of high-resolution time series data provides a comprehensive analysis which would be unattainable by using one clustering method alone.

4.5.3 Validation of results

Expression levels of a number of transcripts were validated by comparing with the results of a previous experiment where RNAseq analysis of GD33DH responding to a higher level of salt shock (500 mM) was sampled at 1 hpt. Several marker genes, which have been previously implicated in the response to high salinity, whose expression overlapped with differentially expressed transcripts in the time-course experiment were identified mapping to Bo9g014980.1 (*MYB96*), Bo3g066190.1 (*NCED3*) and Bo1g007700.1 (*ABF3*). That these key genes were up-regulated across both the time series experiment and the RNAseq validation experiment supports the integrity of the time series data. Despite seeing up-regulation of these key genes, there was only an overlap of 222 genes between both experiments, likely as a result of the differences in experimental conditions, and that different technologies were used to measure gene expression. To fully validate the time series experiment, further validation using qPCR would be desirable.

4.5.4 Limitations of array and experimental design

The time-course microarray experiment does present certain limitations that must be taken into consideration when analysing the results. First, not all of the TO1000 genome was represented on the array, 53,389 Bo genes out of 59,225 Bo genes (90.1%). Due to the highly repetitive nature of the *B. oleracea* genome it was not possible to make unique probes for all genes, so these were not included in the final design. The methods used to expand the information in current TO1000 transcriptome included using transcripts from an RNAseq experiment of 1 hpt salt shock and 24 hpt of cold stress, therefore there is likely to be a bias on the array of additional transcripts related to these conditions, and fewer transcripts relating to late salt shock. The genes represented as transcripts across the array were not present in equal numbers, ranging from one transcript per gene to over 10 probes per gene. In the analysis, the greatest care was taken to remove duplicates where possible.

Secondly, the method of sampling must be considered as it was whole leaf, mixed cell types by nature. This means that some of the more specific effects of individual cells that respond in a unique manner e.g. stomatal cells, would be diluted out by the mix of different cell types present. Should the response of these cells types be of interest, it would be necessary to repeat the experiment with the cells of interest tagged with a fluorophore and carry out Fluorescence-Activated Cell Sorting (FACS) to separate the cells of interest from the general cell population and carry out gene expression analysis on these individually, however this was beyond the scope of this experiment. The final limitation was the short time frame in which sampling occurred. It would be interesting to see the transcriptomic changes involved in salt shock beyond 36 h, however again, this was beyond the scope of this experiment.

4.5.5 Further work

There were several aspects of this experiment that would, given more time, be explored. These included:

- A more in-depth model of the gene regulatory networks in GD33DH responding to salt shock. A solution to the strong circadian presence in the data would be necessary in developing a stress model in GD33DH, which was beyond the scope of this thesis. Advantages of carrying out high resolution time series analyses is the ability to model gene regulatory network to establish key regulatory transcripts (Penfold and Buchanan-Wollaston, 2014) and would be of great interest in further establishing key regulators and potential breeding targets for *B. oleracea*.
- Further biochemical, metabolic and physiological measurements of GD33DH in

response to salt shock such as photosynthetic rate, proline content and yield would be useful in strengthening the results of the transcriptomic analysis.

- As previously mentioned, more validation to strengthen the results of the microarray analysis.
- A large number of novel differentially expressed transcripts were identified in GD33DH responding to salt shock, it would be interesting to investigate these transcripts further to see if novel functions and potential genes of interest were present.

4.6 Chapter Summary

This chapter has focused on the generation and analysis of a high resolution time series experiment to determine the effects of salt shock on the *B. oleracea* GD33DH transcriptome over a period of 36 hpt. An extensive change in gene expression takes place over the time period whilst the plant acts to protect itself from the effects of increasing Na^+ ions in the leaves. Metabolism is altered to reduce the effects of the stress to enable continuation of normal growth. Cross talk between multiple hormone signalling pathways, predominantly ABA, JA and ethylene initiate the differential expression of a vast number of TFs, which are involved with the co-ordination this drastic reprogramming of the transcriptome.

The results of this experiment greatly enrich the existing information on potential salt tolerant mechanisms of *B. oleracea* and provide numerous candidate genes for further analyses to improve the salt tolerance of *Brassica* crops.

Chapter 5

Analysis of gene expression in response to salt shock in wild C-genome *Brassica* species

5.1 Chapter overview

In the previous chapter, the response to salt shock was thoroughly investigated in *B. oleracea* GD33DH, a DH broccoli developed from the commercial line ‘Green Duke’. GD33DH has been used as a parent of several well studied mapping populations including the A12 x GD33 (AG) population (Bohuon et al., 1996, 1998; Broadley et al., 2008; Issa et al., 2013) and the Mar34 x GD33 (MGDH) population (Walley et al., 2011). The investigation was carried out using a high resolution time-series analysis through which many differentially expressed genes, including regulatory genes were identified. This analysis was used to infer the early biological mechanisms that GD33DH uses to limit the damage caused by salt shock.

Commercially developed lines may lack genetic diversity due to selective breeding, resulting in a narrow gene pool which may not be capable of responding to unfavourable environments (Reeves et al., 2012). Crop wild relatives are species that are closely related to cultivated varieties. These may possess the characteristics that are necessary for the adaption of crop plants to harsher, less predictable environments in order to secure global food production for future generations (Dempewolf et al., 2014). Crop wild relatives have been identified as an important but often neglected resource of genetic material that provide breeders with a wider gene pool from which to draw allelic diversity in the development of improved crop species.

Much work has been carried out to develop genetically diverse, pre-breeding material for *B. oleracea*, consisting of germplasm of wild relatives of C-genome *Brassica* species (Pink et al., 2008; Walley et al., 2012), known as the Cg-Diversity Fixed Foundation Set

(Cg-DFFS) as described in Chapter 1. It is an ideal collection to explore the genetic diversity involved with the response to salt shock and to identify associated genes and pathways. As the collection comprising of DH lines generated from wild *Brassica* species it is ideal for analysis by RNAseq as the reads produced can be easily aligned to the genome without issues associated with phase. Also, once interesting genes have been identified, the ability for further study and introgression to commercial breeding lines is made easier by the reproducibility of the fixed genetic component of the lines between generations, in addition to the rapid and reliable development of the lines to the reproductive stage which is not necessarily the case when using the founder lines as discussed in Chapter 1.

The aim of this chapter was to investigate the natural variation in tolerance to salt shock in C-genome wild *Brassica* species. In order to determine the effects of salt shock on the transcriptomes of wild *Brassica* species, a small collection of genetically fixed, doubled haploid species underwent RNAseq analysis to investigate differences in gene expression following salt shock. This involved the following the following steps:

1. An exploration of the natural variation in growth of C-genome wild *Brassica* species following salt shock.
2. Transcriptome (RNAseq) analysis of selected Doubled Haploid *Brassica* lines.
3. The identification and functional analysis of the differentially expressed genes and comparisons between tolerant and susceptible lines.

The value of this analysis is to determine salt tolerance mechanisms in wild C-genome *Brassica* species and to identify suitable breeding material to introgress into *B. oleracea* for the future development of stress tolerant varieties.

5.2 Results

5.2.1 Preliminary salt shock phenotype screen of wild C-genome *Brassica* species

A selection of S1 lines from the Cg-DFFS collection (selfed from founder lines; Figure 5.1) were subject to a salt shock and then screened for various phenotypic traits (plant height, leaf area and dry weight) to investigate the extent of variation in the response to salt shock. The screen resulted in the identification of a number of C-genome *Brassica* lines exhibiting higher levels of tolerance or sensitivity compared to the rest of the population.

The species used in the preliminary screen contained a selection of cultivated *B. oleracea* lines as well as wild C-genome S1 lines, such as wild *B. incana*, *B. cretica*, *B. macrocarpa* and *B. montanta* (Figure 5.1). To carry out the preliminary screen, a salt

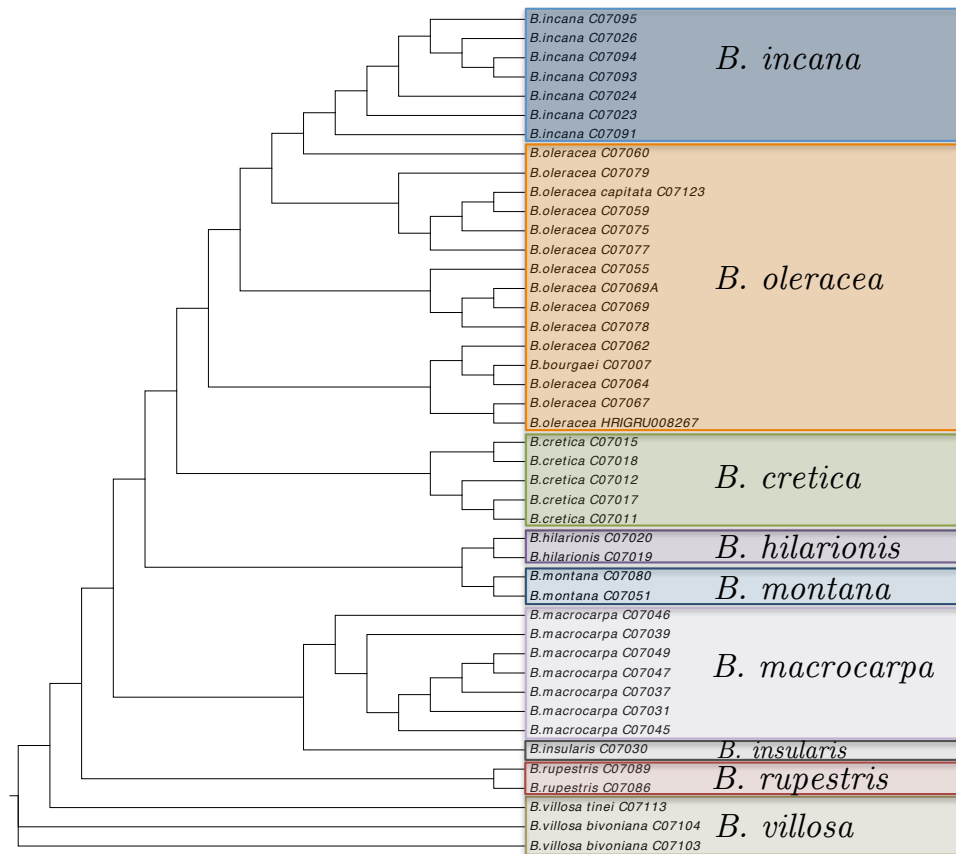


Figure 5.1: Phylogenetic tree of wild C-genome *Brassica* species (S1 lines) used in the study

This figure shows a phylogenetic tree produced from genomic data of the S1 lines used in the initial stage of this study. The species associated with each line is given in a coloured box as indicated in the figure. The tree was produced using phyML based on the maximum likelihood principle (Guindon et al., 2009).



(a) *Brassica macrocarpa* C07039



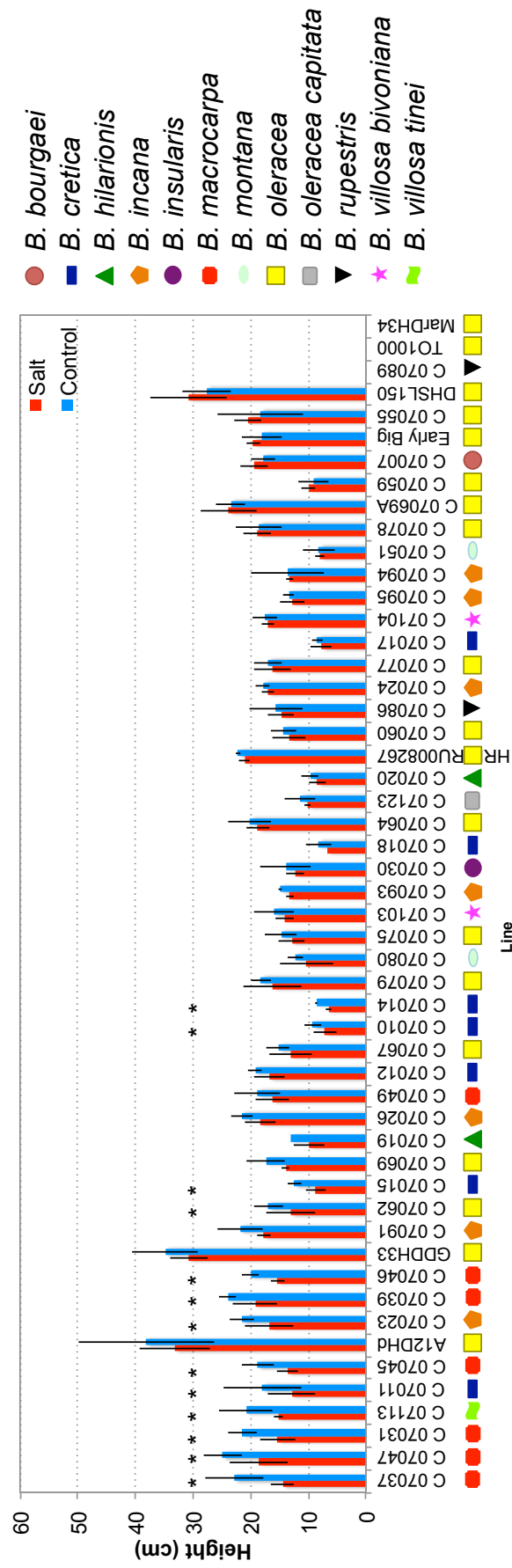
(b) *Brassica bourgaei* C07007

Figure 5.2: Effect of salt shock on selected wild C-genome *Brassica* species

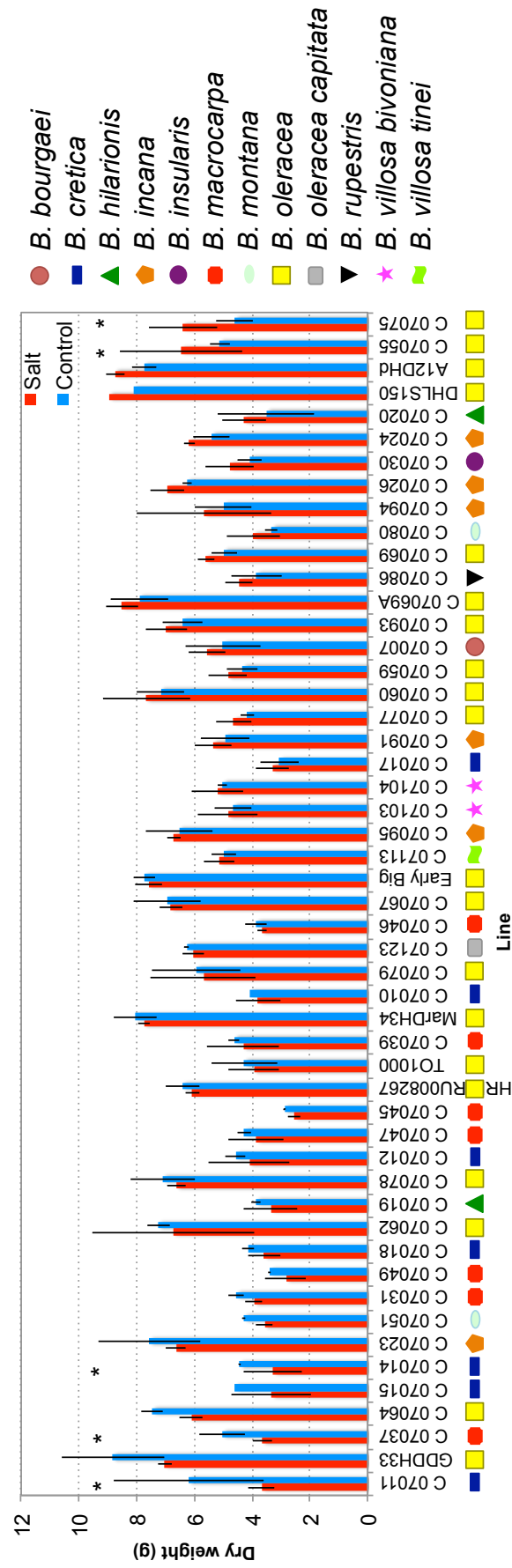
Wild C-genome *B. macrocarpa* (a) showing susceptibility and *B. bourgaei* (b) showing tolerance to salt shock, 14 days after receiving treatment. Scale bar is 5 cm.

shock was applied to these 50 different Cg-DFS S1 lines ($n=3$), as previously described in Methods (see Chapter 2). A period of normal watering was resumed for 14 days after treatment to let the plants recover with the aim of testing resilience to salt shock, rather than the initial stress response. After the recovery period of 14 days, several plant traits were measured to assess the effect of the salt shock on different physiological aspects of growth.

Following measurement of plant height, dry weight and leaf area, it was clear that there was a large amount of variation in response to salt shock in the wild *Brassica* species with respect to the measured traits. Some plants showed a susceptible phenotype, such as *B. macrocarpa* C07039 (Fig. 5.2a), whilst others showed a tolerant phenotype such as *B. bourgaei* C07007 where the salt appeared to have little effect on the visible phenotype (Fig. 5.2b). This indicated potential genetic variation that could be useful for breeding plants with enhanced tolerance to abiotic stress conditions such as salt shock. It should be noted that despite an attempt to remain consistent in the photographing of plants for height measurements, it is possible that there may be some distortion of distance, therefore there will be a minor amount of inaccuracy associated with this measurement.



(a) Plant height



(b) Whole plant dry weight

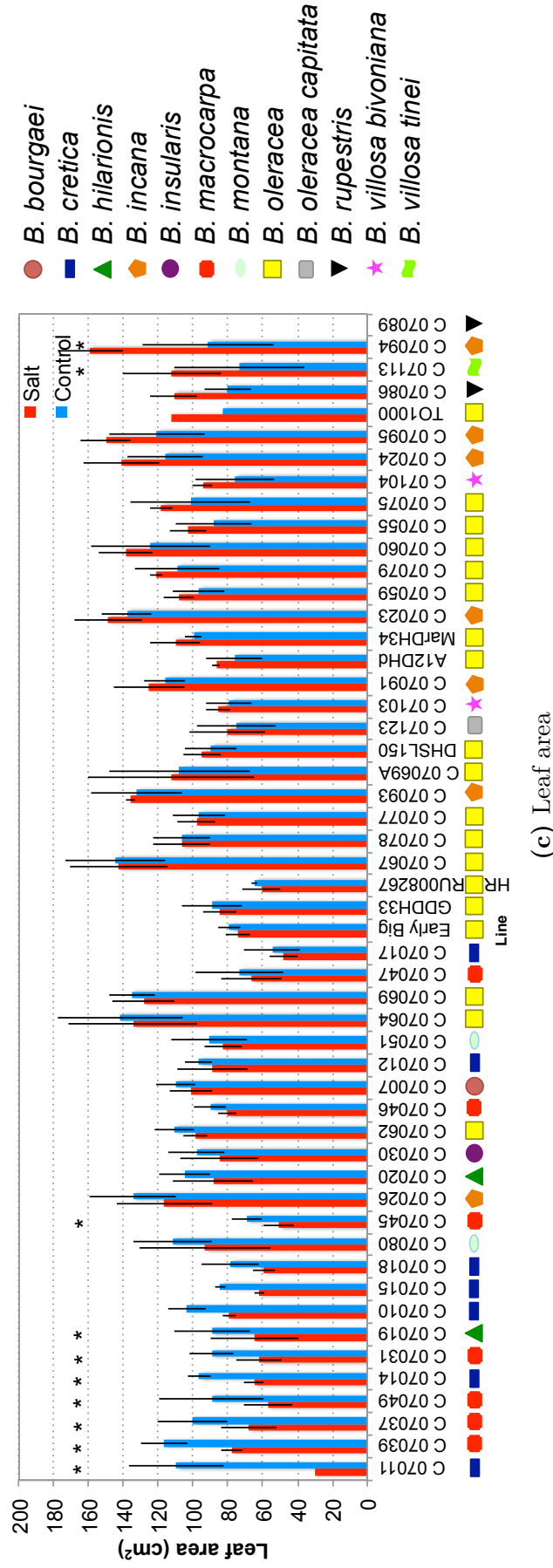


Figure 5.3: Natural variation of resilience to salt shock in wild C-genome *Brassica* species (S1 lines)

The difference in natural variation in response to salt shock in wild C-genome *Brassica* species in treatment verses control. Responses measured include plant height (a), whole plant dry weight (b) and leaf area (leaf #5; c) of plants 14 days after receiving salt shock treatment. Significance, determined by LSD at 0.05 ($n=3$), is indicated by an asterisk (*). The x -axis shows the measurement, the y -axis shows the line and corresponding species, as per the key. Treatment is shown by the red bars, control by the blue. The error bars show the standard deviation from the mean.

Figure 5.3a shows the plant height measurements for each line under salt conditions. There was a large degree of variation between the lines, with average difference between control and treatment ranging from between -14.4% to +5.7%. The majority of species showed a decrease in height following salt shock and this was statistically significant in several *B. cretica* and *B. macrocarpa* lines, as indicated in Figure 5.3a. *B. oleracea* species generally performed the most effectively under salt shock, showing no decrease in height following salt shock, for instance the cultivated *B. oleracea* DHSL150 and Early Big and wild lines such as C07059 and C07055 (both wild *B. oleracea*).

In terms of whole plant dry weight (Fig. 5.3b), there was also variation present within the lines, with the average difference between lines ranging from -34.7% to 23.2%. As a general trend, species which have a greater evolutionary distance from cultivated *B. oleracea* such as *B. cretica* and *B. macrocarpa* showed the greatest decrease of dry weight in response to salt shock. There were a small number of lines found at the two extremes either showing a significant increase in dry weight compared to the controls eg wild *B. oleracea* lines C07055 and C07075, or a significant decrease in dry weight compared to control e.g. lines C07011, C07007 and C07014. The literature reports both increases in dry weight in plants following salt stress (Andriolo et al., 2005; Qados, 2011) and also decreases in dry weight (Jamil et al., 2007; Memon et al., 2010). Gain of dry weight in plants following salt shock could be caused by an increase in waxy deposits on the surface of the leaf under water stress conditions to prevent further water loss (Kosma et al., 2009). A decrease in plant dry weight may be caused by a drop in metabolism resulting in reduced growth.

Variation was also seen with regards to leaf area (leaf #5), with the average difference between lines ranging from -29.7% to 13.8% (Fig. 5.3c). There were two lines which showed a significant increase in leaf area following salt shock, C07094 (*B. incana*) and C07113 (*B. villosa tinei*). Several lines showed a significant decrease in leaf area following salt shock including C07011 (*Brassica cretica*), which showed a particularly large reduction in leaf area in response to salt shock. Also several *B. macrocarpa* lines and a *B. hilarionis* lines showed reduced leaf area under high-salt conditions. A reduction of transcripts involved in cell expansion and leaf development following salt shock were seen in GD33DH (Fig. 4.9) as a result of Ca^{2+} signalling following perception of elevated Na^{+} levels in the soil. In rice and bean plant (*Vicia faba*), an effect on leaf area following salt stress has also been observed (Ali et al., 2004; Qados, 2011). Generally, plant height has been considered the most useful 'rough' estimation of tolerance level of *Brassica* plants to salt stress conditions (Su et al., 2013) as plant growth has been widely shown to be affected by salinity.

5.2.2 Differential responses to salt shock in wild C-genome *Brassica*

In an attempt to classify the responses of the different wild S1 *Brassica* lines to salt shock, the resilience measurements (Fig. 5.3) were clustered to produce 8 clusters using *K*-means clustering methods (Fig. 5.4). These clusters group together lines with similar differences in their Salt Tolerance Indices and could be used to suggest different or overlapping response mechanisms in wild S1 *Brassica* lines. Cluster 1 consisted of lines that were the most susceptible to salt shock overall, showing statistically significant decreases height, leaf area and dry weight. The group consisted of three *B. cretica* lines, three *B. macrocarpa* lines and one line of *B. hilarionis*. The second cluster showed a mostly statistically significant negative effect on plant height, but the effect on leaf area and plant dry weight was less significant compared to the first cluster. This group consisted of three lines of *B. cretica*, three lines of *B. macrocarpa*, three lines of *B. oleracea* as well as one *B. incana*. Cluster 3 showed no effect on plant dry weight following salt shock but there were negative changes in height and leaf area whilst Cluster 4 showed almost the opposite with a negative effect on plant dry weight, whilst plant height and leaf area were neutrally affected by the stress. Cluster 5 and 6 only showed susceptibility in terms of decreased height, whilst leaf area and dry weight showed a neutral response to salt shock. Clusters 7 and 8 contained the lines which were most tolerant to the salt shock and showed a neutral effect in all three traits. These two clusters contained four lines of *B. oleracea*, two lines of *B. incana* and one line each of both wild *B. villosa bivonia* and *B. bourgaei* suggesting that some of these wild species of C-genome *Brassica* will be suitable to contribute genetic material for increasing abiotic stress tolerance in cultivated *B. oleracea* crops.

Based on this analysis, several lines were selected for the next stage of the analysis and these included (in order of increasing tolerance) C07019 (Cluster 1; *B. hilarionis*), C06079 (Cluster 2; *B. oleracea*), C07069 (Cluster 3; *B. oleracea*), C07060 (Cluster 5; *B. oleracea*), C07007 (Cluster 7, *B. bourgaei*) and C07094 (Cluster 8; *B. incana*).

5.2.3 Second salt resilience screen with Doubled Haploid C-genome wild *Brassica* lines with selection of lines for sequencing

The use of natural variation for crop improvement is one of the main principles of plant breeding, however it is often the case that the specific genes and biological mechanisms of the desired trait are not known (Flowers et al., 1997). Mapping of Quantitative Trait Loci (QTL) under pinning biological traits relies on the knowledge of phenotype and genotype however, until recently, genotyping was often carried out using low throughput, highly

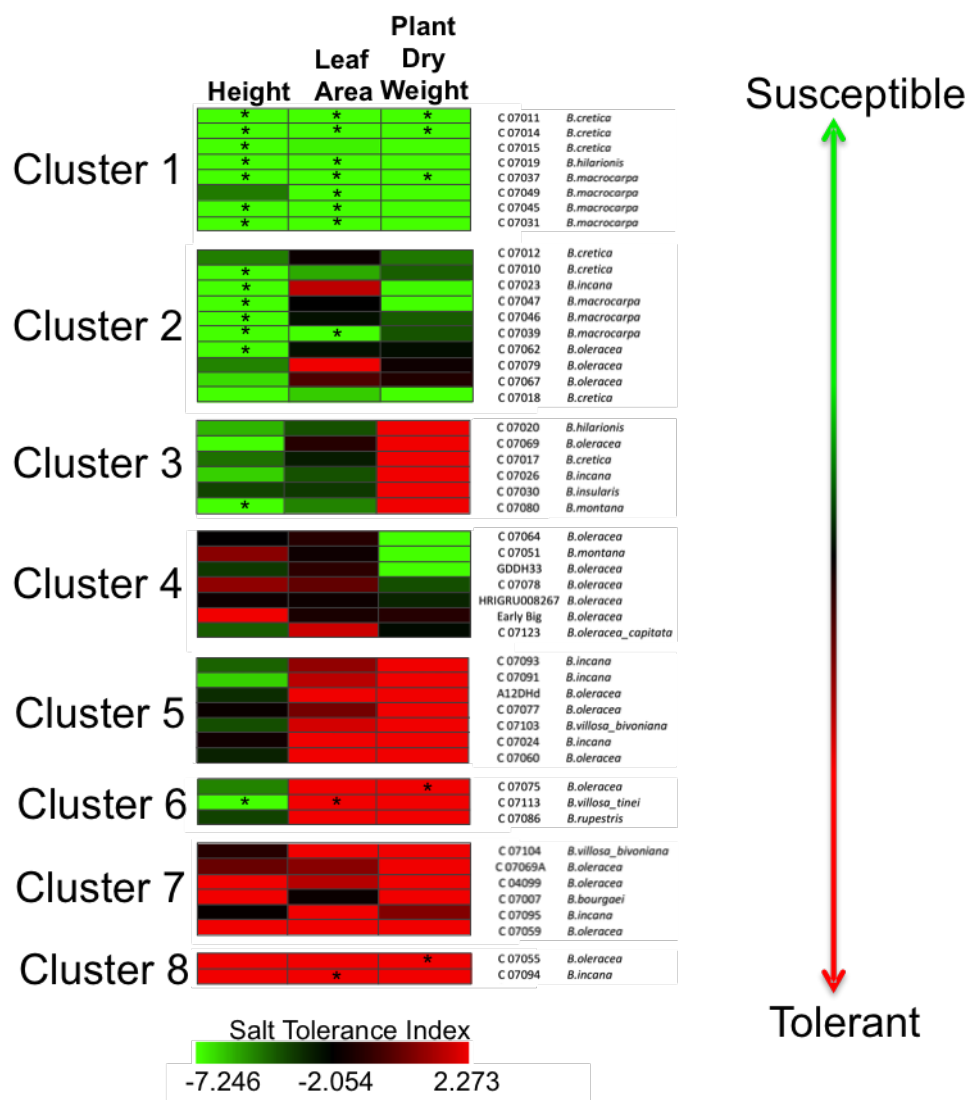


Figure 5.4: Clustering wild C-genome *Brassica* species based on response to salt shock

K-means clustering of average differences (control - treatment) for plant height, leaf area and whole plant dry weight. Green shows a negative (ie decrease) effect of the stress on the trait, black shows the midpoint of the scale (low negative) and red shows a positive effect (ie an increase of the trait following salt shock), as per the scale. Statistically significant measures are marked with an asterisk (*)

Line Number	Founder Line	Species	Line Type
C07007	C04006	<i>B. bourgaei</i>	S1
C07019	C04015	<i>B. hilarionis</i>	S1
C07060	C04052	<i>B. oleracea</i>	S1
C07069	C04062	<i>B. oleracea</i>	S1
C07079A	C04069	<i>B. oleracea</i>	S1
C07094	C04081	<i>B. incana</i>	S1
C10025	C04006	<i>B. bourgaei</i>	DH
C10027	C04006	<i>B. bourgaei</i>	DH
C10121	C04069	<i>B. oleracea</i>	DH
C10125	C04052	<i>B. oleracea</i>	DH
C10128	C04052	<i>B. oleracea</i>	DH
C10132	C04062	<i>B. oleracea</i>	DH
C10139	C04062	<i>B. oleracea</i>	DH
C13001	C04015	<i>B. hilarionis</i>	DH
C13012	C04081	<i>B. incana</i>	DH
C13013	C04081	<i>B. incana</i>	DH
DHSL150	-	<i>B. oleracea</i>	Cultivated
Early Big	-	<i>B. oleracea</i>	Cultivated
TO1000	-	<i>B. oleracea</i>	Cultivated

Table 5.1: Summary of Cg-DFFS lines used in the second diversity study

The lines used in the second diversity study, along with the founder line from which it originated, the species and finally the type of line (refer to Figure 1.5 for description on the different line types).

spaced molecular markers such as SSRs (Quesada et al., 2002). Advances in biotechnology resulting in cheaper and more rapid sequencing have enabled the development of different methods for identifying genes involved in beneficial traits such as tolerance to abiotic stress conditions. The transcriptomes of a selection of lines with differing tolerance to salt shock were sequenced in order to gain information on gene expression with the aim of investigating salt tolerance in C-genome *Brassica* species.

5.2.4 Experimental design

A selection of DH lines from the Cg-DFFS collection, generated from the selected lines, were screened for their response to salt shock (Table 5.1). Three of these lines were used for RNAseq analysis to compare and contrast gene expression in DH lines which have different resilience to salt shock. The founder lines of each S1 line of interest was traced and two of the respective DH lines were sown (except in the case of founder lines C04069 and C04015 where two DH lines were not available so only one was used). Two DH lines were sown, as each will have a different combination of the parental alleles and thus may exhibit differing levels of tolerance. In additional, the relevant S1 and cultivated *B. oleracea* lines were sown (Table 5.1). The resilience screen was repeated, as described previously. The

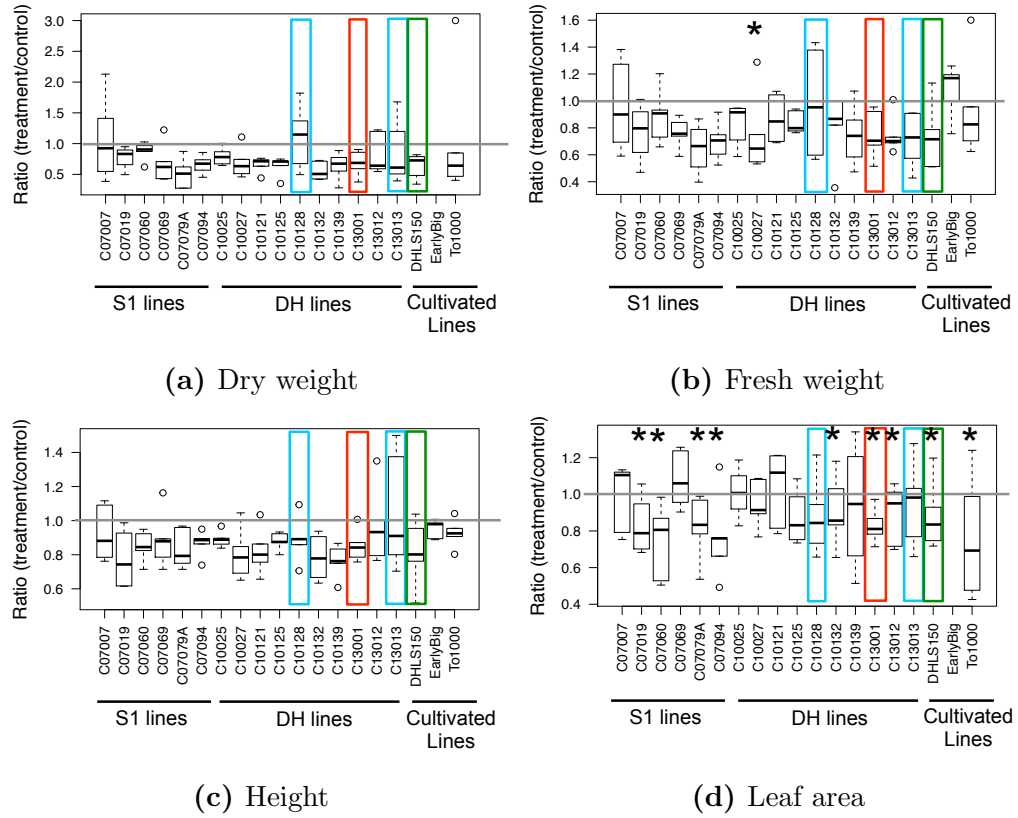


Figure 5.5: Effect of salt shock on selected wild C-genome *Brassica* S1, DH and cultivates lines

Selected S1, DH and cultivated lines in response to salt shock ($n=6$). Measured traits were (a) whole plant dry weight, (b) whole plant fresh weight, (c) plant height and (d) leaf area. The grey line represents a ratio of 1, in which there is no difference between treatment and control samples. The highlighted lines were sent for RNA sequencing. Lines marked in blue were classed as ‘tolerant’ and those in red were classed as ‘susceptible’. DHLS150 (green), is one of the parent lines used in the crosses to genetically fix the DH material. Lines marked with an asterisk (*) indicate a significant difference between treatment and control samples ($p < 0.05$).

resilience screen and data collection were carried out by Almustapha Lawal (University of Warwick, UK) and the results can be seen in Figure 5.5.

The salt shock treatment resulted in a negative effect on all of the lines that were screened but these were significant in only a few samples (Fig. 5.5) and only for the leaf area measurement (except for C10027 which showed a significant difference in fresh weight in treatment compared to control samples). Lines which show tolerance to salt shock will have a smaller difference between treatment and control samples and as a result may show no significant difference following treatment. In terms of dry weight, the tolerant DH lines C10128 (*B. oleracea*), C13012 and C13013 (both *B. incana*) showed the greatest variation within the data, with some of the lines having a mean ratio greater than 1 indicating that no dry weight was lost following salt shock. Other lines decreased in dry weight following salt shock and some showed little difference. This pattern was not necessarily seen when fresh weight was considered, except in the case of C10128 which had the largest range and the highest mean ratio of the DH lines. Most of the lines showed a decrease in fresh weight, which was significant only in the C10027 line following salt shock.

In terms of plant height, C13013 (*B. incana*) was the only line which showed a neutral effect of salt shock on plant growth, indicating that no height was lost following salt shock in some samples though the variation was large. In the rest of the lines a negative effect on height, with smaller amounts of variation was seen.

The final trait to be discussed is leaf area, which showed more interesting and significant results than other traits. A clear decrease in leaf area was seen in most of the samples, which was significant in four S1 lines, three DH lines and two of the cultivated lines.

In addition to allelic variation, variation can be caused either by additive effects such as the combined effect of multiple alleles or by non-additive effects involving the interaction of genes from different genetic backgrounds (epistasis). Such variation in response to salt shock was seen in the S1 lines, suggesting that the genes originating from the wild species, and the interaction of these genes with the DHLS150 genes were having a substantial effect on the response to salt shock. For example, the C13013 line which showed one of the most tolerant responses to salt shock in the experiment. The S1 line related to this line was found in the most tolerant cluster, Cluster 8. In addition, C13001 which was one of the more susceptible lines in this experiment was related to the C07019 S1 line which was found in the most susceptible cluster, Cluster 1 (Fig. 5.4).

To summarise the outcome of this analysis, salt shock has a significantly negative effect on leaf area, and results in a general decrease in plant height and fresh/dry weight. This information was used in the selection of susceptible and tolerant lines for further study by RNAseq as discussed below.

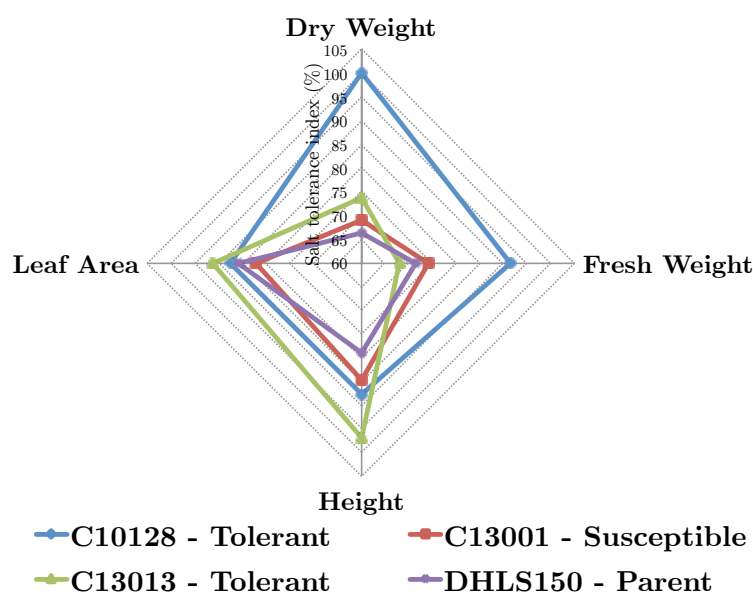


Figure 5.6: Resilience measurements of lines selected for RNAseq

Spider plot showing average difference (%) for each trait, measuring the effect salt shock treated plants compared to control

5.2.5 Selection of susceptible and tolerant DH lines

The outcome of this second screen of S1 and DH lines was the selection of three DH lines with contrasting responses - two tolerant lines (C10128 and C13013) and one sensitive line (C13001) for sequencing, alongside the DHSL150 parent line (Table 5.2).

DH Line	Species	Founder Line	Tolerance
C10128	<i>B. oleracea</i>	C04052	Tolerant
C13013	<i>B. incana</i>	C04081	Tolerant
C13001	<i>B. hilarionis</i>	C04015	Susceptible
DHSL150	<i>B. oleracea</i>	Parent	Susceptible

Table 5.2: Doubled Haploid lines selected for RNAseq

Details of the DH lines used for RNAseq analysis, including information relating to the species from which the lines belong, the founder line used to make the cross, and the susceptibility of the DH line to salt shock.

The phenotype data for these selected lines are summarized in Figure 5.6, where it can be seen that C10128 most successfully preserves its dry/fresh weight following salt shock despite showing a substantial decrease in height. In order to determine whether the tolerance level has been retained between the S1 lines and the resultant DH lines, the output of the second screen (DH lines) was compared to the first screen (S1 lines; Fig. 5.3 and 5.4). The S1 line originating from the same founder line as C10128 is C07060 (Table 5.1) which performed well in all traits apart from leaf area in the second screen (Fig. 5.5)

and in the trait cluster analysis was placed in Cluster 5, which is towards the tolerant end of the clustering, showing a decrease in height and maintenance of dry weight and leaf area (Fig. 5.4). C13013 was also tolerant in the second screen, the S1 originating from the same founder line as C13013 was C07094, found in Cluster 8 in the initial screen, the most tolerant cluster with a significant increase in leaf area following salt shock. Based on the results of the secondary screen, the resilience is preserved in this line following the crossing and microspore culture of C07094 to DHSL150 to generate the DH lines. Finally, C13001 originated from the same founder line as the S1 line C07019 that was found in Cluster 1, the most susceptible cluster with large decreases in all traits, and this is consistent with the DH line screen (see Fig. 5.4). These results show that there was a high level of consistency between the S1 lines and the DH lines relating to their appropriate founder line, indicating that crossing the founder line to DHSL150 and generating DH has not resulted in the loss of genetic variation in salt shock tolerance. Thus a proportion of genetic variation from the wild species that may be affecting tolerance to salt shock has been preserved in the DH lines used further in the study.

5.2.6 Mineral analysis of selected tolerant and susceptible DH lines

Mineral analysis was carried out on two of the samples selected for sequencing, C10128 (tolerant) and C13001 (susceptible) lines using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) with freeze dried leaf material (sampled at 24 hpt) to determine levels of Na^+ and K^+ in the leaf. The ICP mass spectrometry was carried out by Almustapha Lawal (University of Warwick) and the results are shown in Figure 5.7. Whilst the Na^+ ions stay at a similar level in both lines, C10128 has a vastly increased level of K^+ ions compared to C13001. This resulted in a higher $\text{K}^+:\text{Na}^+$ ratio for the tolerant line C10128 ($\text{K}^+:\text{Na}^+$ ratio=6.11) compared to the susceptible line C13001 ($\text{K}^+:\text{Na}^+$ ratio=2.54). A high $\text{K}^+:\text{Na}^+$ ratio due to the retention of K^+ ions has been correlated with increased tolerance to high salt environments in *Arabidopsis* (Maathuis and Amtmann, 1999; Sun et al., 2015), which is consistent with the results seen in this experiment. K^+ deficiency causes a sharp decrease in tolerance in maize under salt stress, K^+ deficiency resulted in a reduced ability to assimilate nitrogen and photosynthetic carbon, whilst also affecting the light reaction pathways of PSI and PSII (Qu et al., 2012, 2011). Potassium also has a role in the osmotic adjustment and maintaining turgor pressure in plants under salt stress conditions (Munns and Tester, 2008). High levels of potassium have been shown to reduce oxidative damage caused by ROS in the cytosol of maize seedlings (Gong et al., 2010). These results suggest that the ability of *Brassica* lines to accumulate high levels of K^+ in the shoot following salt shock may reflect the tolerance level of the plant. Given the

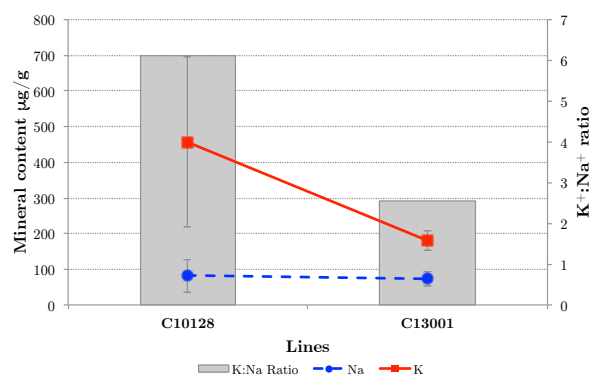


Figure 5.7: Mineral analysis of a tolerant and a susceptible Doubled Haploid *Brassica* line

Plot shows results of ICP mineral analysis (carried out by Almustapha Lawal, University of Warwick) for one tolerant line (C10128, *B. oleracea*) and one susceptible line (C13001, *B. hilarionis*). The blue dashed line is read off the primary axis (left) and indicates Na⁺ mineral content of freeze dried leaf material sampled 24 hpt. The red line is read off the primary axis (left) and indicates the K⁺ mineral content. The Na⁺/K⁺ ratio is indicated by the grey bars on the secondary axis (right). Error bars represent the standard deviation.

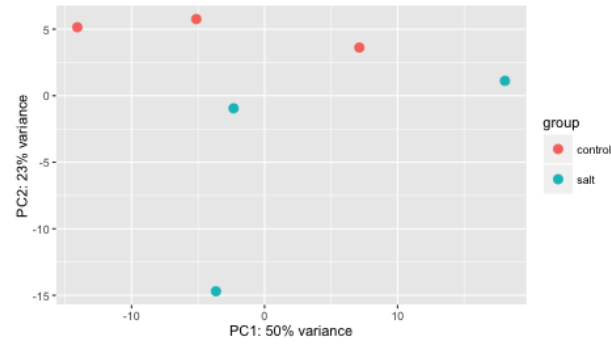
small sample size of only two lines, these results are not conclusive until a more extensive analysis has been conducted.

5.2.7 Transcriptomic analysis of selected tolerant and susceptible lines

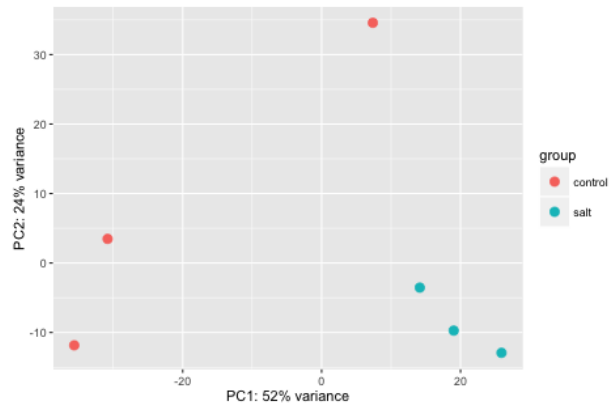
Analysis of the differences in gene expression following salt treatment between a few wild *Brassica* lines showing contrasting salt tolerance, may enable insight into the genetic and physiological responses underpinning salt shock tolerance. A similar approach has been carried out in a variety of plant species including barley (Gao et al., 2013; Guo et al., 2009), rice (Jiang et al., 2013a) and Arabidopsis (Wang et al., 2013).

Three DH lines and the DHSL150 parent involved in the development of genetically stable lines from wild *Brassica* species, with contrasting tolerance levels, were selected for transcriptome sequencing, as discussed above (Table 5.2). Salt treated and control leaf material for each line selected above was collected 24 hpt ($n=3$), snap frozen in liquid nitrogen followed by high quality RNA extractions. Whole transcriptome sequencing (RNAseq) was carried out on the material in order to gain knowledge on transcript expression under the given conditions. This time point was chosen because in the time-series experiment (described in Chapter 4) 24 hpt was shown to be the time at which the greatest number of differentially expressed genes were detected in response to salt shock in GD33DH (Fig. 4.8b).

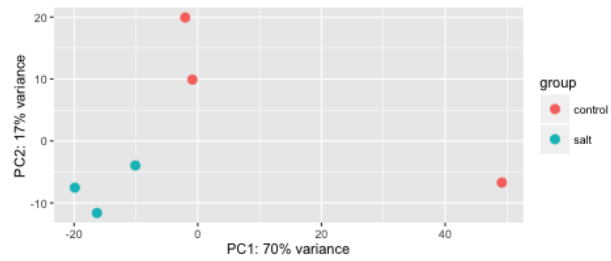
The selected lines were sequenced using the Illumina HiSeq 2000 to produce paired end reads of 100bp. The quality of the reads was assessed using FastQC and found to be



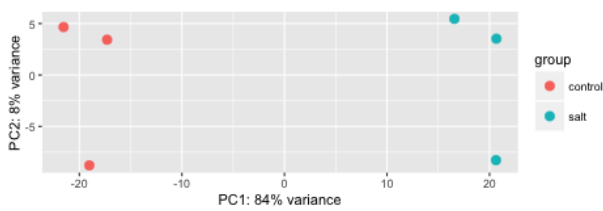
(a) C10128



(b) C13013



(c) C13001



(d) DHSL150

Figure 5.8: PCA loading plots of the RNAseq count data

Principle Component Analysis (PCA) performed using DESeq2 log-normalised RNAseq count data. Loadings for the first principle component (PC1) and PC2 are located on the *x*-axis and *y*-axis, respectively. The samples were (a) C10128, (b) C13013, (c) C13001 and (d) DHSL150.

Line	Species	Total DE	Up	Down
C10128	<i>B. oleracea</i>	48 (4)	10 (2)	38 (2)
C13013	<i>B. incana</i>	5,479 (409)	2,413 (221)	3,066 (188)
C13001	<i>B. hilarionis</i>	1,853 (180)	597 (59)	1,256 (121)
DHSL150	<i>B. oleracea</i>	4,310 (312)	1,763 (178)	2,547 (134)

Table 5.3: Differentially expressed genes in four DH *Brassica* lines following salt shock treatment

Testing for differential expression was carried out in DESeq2 using an adjusted P-value < 0.05 and a $\logFC > 1$ as threshold of significance. Genes with a positive \logFC were considered up-regulated and those with a negative \logFC were down-regulated. The number of transcription factors are given in parenthesis.

outstanding (Phred score > 30) and as such no pre-processing of the reads was carried out (Appendix H). The reads were aligned to the *B. oleracea* TO1000 genome (Parkin et al., 2014) and counted, as previously described (Chapter 2). A PCA plot of the log-normalised count data of each line indicated that there several of the samples were outliers in the data, see Figure 5.8. No clear clustering was seen for C10128 (Fig. 5.8a) suggesting that there is not a great amount of difference between salt-treated and control samples. Both C13013 and C13001 (Fig. 5.8b and c) showed clustering of the treatment and control samples, but each with a control sample as an outlier, which was removed before subsequent analysis. DHSL150 (Fig. 5.8d) showed clear clustering of treatment and control and therefore no samples were removed prior to analysis.

Once the outlying samples were removed, differential expression analysis comparing treated samples to control was carried out using DESeq2. DESeq2 models count data based on the negative binomial distribution, from which parameters such as the mean and dispersion are measured from the data (Love et al., 2014). A threshold was set for differential expression in which genes must have an $FDR < 0.05$ (adjusted p -value) and a $\logFC > 1$ (Log ratio of treatment/control) to be considered differentially expressed, as shown in Table 5.3.

The number of differentially expressed genes varied greatly between lines, from 48 (C10128) to 5,479 (C13013) across the experiment, suggesting that each of the lines may employ different mechanisms of salt tolerance. The results show that there were more down-regulated genes in all lines compared to up-regulated genes, similar to the results seen in the GD33DH time-course experiment (Chapter 4). The list of differentially expressed genes for each sequenced line is made available in additional datafile4 (Appendix I).

Bo gene ID	Bo description	C10128	C13001	C13013	DHSL150
Bo7g081850.1	ABC transporter G family member	1.543	-	1.462	1.541
Bo4g098210.1	tRNA dimethylallyltransferase	1.407	-	-	-
Bo6g005990.1	Vacuolar iron transporter-like protein	1.339	-	3.094	-
Bo7g012980.1	Histidine kinase	1.168	-	-	-
Bo4g115480.1	Proline transporter	1.164	-	-	-
Bo3g001360.1	Ferritin	1.065	2.013	3.546	1.093
Bo8g098440.1	tRNA dimethylallyltransferase	1.059	-	-	-
Bo6g062410.1	Homeobox-leucine zipper protein family	1.020	2.376	2.465	-
Bo1g006740.1	Catalase	1.018	-	1.706	-
Bo6g018000.1	Eukaryotic translation initiation factor 5A-1	1.006	2.231	2.512	2.560
Bo8g049770.1	Abcscic acid receptor PYR1	-1.014	-2.014	-2.586	-2.751
Bo9g117290.1	protein kinase family protein	-1.015	-3.143	-3.644	-3.382
Bo6g035440.1	germin-like protein	-1.024	-2.856	-3.287	-3.515
Bo3g010840.1	conserved hypothetical protein	-1.032	-3.100	-3.572	-3.683
Bo2g043370.1	conserved hypothetical protein	-1.038	-2.014	-2.261	-3.058
Bo3g036020.1	bZIP transcription factor family protein	-1.038	-3.676	-3.612	-4.614
Bo4g165190.1	LONELY GUY (LOG)	-1.041	-	-1.827	-
Bo3g090210.1	Transposon protein, putative, Pong sub-class	-1.059	-	-	-
Bo2g161000.1	serine carboxypeptidase-like protein	-1.076	-2.431	-2.680	-2.354
Bo3g140130.1	Transducin/WD40 repeat-like superfamily protein	-1.098	-	-	-
Bo4g187530.1	Protein of unknown function, DUF538	-1.115	-1.561	-1.581	-1.351
Bo3g175030.1	Cation transport regulator-like protein	-1.117	-	-	-
Bo4g190200.1	Major facilitator superfamily protein	-1.119	-2.436	-2.120	-
Bo3g035060.1	Expansin	-1.130	-3.550	-4.021	-4.617
Bo1g158950.1	Carbonic anhydrase	-1.140	-	-	-
Bo4g108180.1	phy rapidly regulated	-1.171	-3.612	-4.595	-4.891
Bo00916s020.1	Exostosin family protein	-1.173	-2.099	-2.146	-2.850
Bo5g017210.1	Bifunctional inhibitor/lipid-transfer protein/ seed storage 2S albumin superfamily protein	-1.182	-4.624	-4.961	-5.734
Bo3g045130.1	hypothetical protein	-1.195	-	-	-
Bo3g039170.1	arabinogalactan protein	-1.212	-4.765	-5.788	-4.461
Bo3g004570.1	HXXXD-type acyl-transferase-like protein	-1.239	-	-	-1.811
Bo6g075570.1	Arabinogalactan peptide	-1.251	-3.096	-3.268	-4.752
Bo2g049350.1	Phosphate-induced (Phi-1) protein, putative	-1.278	-1.211	-1.219	-1.067
Bo3g079960.1	Bifunctional inhibitor/lipid-transfer protein/ seed storage 2S albumin superfamily protein	-1.279	-2.026	-	-3.266
Bo9g004020.1	basic helix-loop-helix (bHLH) DNA-binding	-1.295	-2.674	-2.931	-3.534
Bo00285s340.1	Aquaporin	-1.316	-2.225	-2.994	-3.060
Bo2g009740.1	myo-inositol-1-phosphate synthase	-1.437	-2.616	-4.007	-3.911
Bo7g110110.1	Plant invertase/pectin methylesterase inhibitor	-1.442	-2.687	-3.117	-3.989
Bo9g022280.1	Bifunctional inhibitor/lipid-transfer protein/ seed storage 2S albumin superfamily protein	-1.511	-5.239	-5.201	-8.339
Bo4g114100.1	conserved hypothetical protein	-1.602	-	-1.063	-
Bo01938s010.1	Glutathione S-transferase T3	-1.749	-2.786	-5.841	-
Bo4g025260.1	Ribosomal protein-like protein	-1.772	-	-	-
Bo8g090790.1	conserved hypothetical protein	-1.800	-	-2.563	-
Bo2g082290.1	hypothetical protein	-1.823	-2.613	-3.066	-4.316
Bo2g012470.1	Gibberellin-regulated protein	-1.855	-1.917	-	-1.705
Bo3g024810.1	conserved hypothetical protein	-2.008	-	-2.498	-
Bo9g177270.1	conserved hypothetical protein	-2.320	-2.672	-3.333	-
Bo4g173400.1	conserved hypothetical protein	-2.399	-	-3.368	-

Table 5.4: Differentially expressed genes in C10128 following salt shock treatment

Testing for differential expression was carried out in DESeq2 using an adjusted P-value < 0.05 and a logFC > 1 as threshold of significance. Genes with a positive logFC were considered up-regulated and those with a negative logFC were down-regulated. Table includes all genes differentially expressed in C10128 and the expression data of genes differentially expressed in the other sequenced lines.

5.2.8 Analysing genes differentially expressed in C10128 responding to salt shock

C10128 has few differentially expressed genes (Table 5.3 and 5.4) suggesting that there are few transcriptomic changes in the leaf following salt shock. Of the 48 differentially expressed genes in C10128, there were a handful of transporter genes including an up-regulated genes encoding for an ABC transporter, a vacuolar iron transporter like protein and a down-regulated aquaporin. This suggest that there is some regulation of ion homeostasis, in which excess Na^+ ions are potentially being stored in the vacuole and down-regulation of an aquaporin assists in the regulation of water loss from the cell (Boursiac et al., 2005). These ion transporters are also differentially expressed in at least one other line, suggesting they are important in the response to salt shock in *Brassica*.

In addition to ion transporters, there was a handful of transcription factors which are differentially regulated in C10128, including up-regulation of a HB and histidine kinase and down-regulation of bZIP and bHLH TFs, suggesting that there is some regulation of stress related gene expression in this line. There is also evidence of hormone regulation, including the down-regulation of Cytokinin riboside 5'-monophosphate phosphoribohydrolase (*LONELY GUY*) involved in the activation of the cytokinin signalling pathway (Kuroha et al., 2009). A down-regulated gibberellin-regulated protein Gibberellic Acid-stimulated Arabidopsis (*GASA4*) was present, which has been shown to have a role in the response to reactive oxygen species and GA responses in Arabidopsis (Rubinovich and Weiss, 2010). Down-regulation of these transcription factors was seen in the other lines sequenced (C13001, C13013 and DHSL150) suggesting that they play important roles in the core response to salt shock. Finally, the ABA receptor *PYR1* was down-regulated across all lines, as was seen in GD33DH (Fig. 4.10).

There was evidence of an effect of salt shock on the lipid membranes, due to the substantial down-regulation of genes relating to lipid transport (three genes encoding bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily proteins) and phospholipid metabolism (myo-inositol-1-phosphate synthase) in C10128 and in the other sequenced lines suggesting this group of genes is incredibly important in the response to salt shock. The down-regulation of an expansin encoding gene, responsible for cell wall loosening was also observed. Altered expression levels of the expansin group of genes have been correlated with shoot growth in maize (Geilfus et al., 2010). The down-regulation of this gene in C10128 may be responsible for the decrease in shoot height seen in this line following salt shock (Fig. 5.6).

The majority of differentially expressed genes in C10128 (totalling 38 genes) were differentially expressed in at least one other line (Table 5.4) and generally show a high logFC (positive or negative) in the same direction suggesting that these genes are important

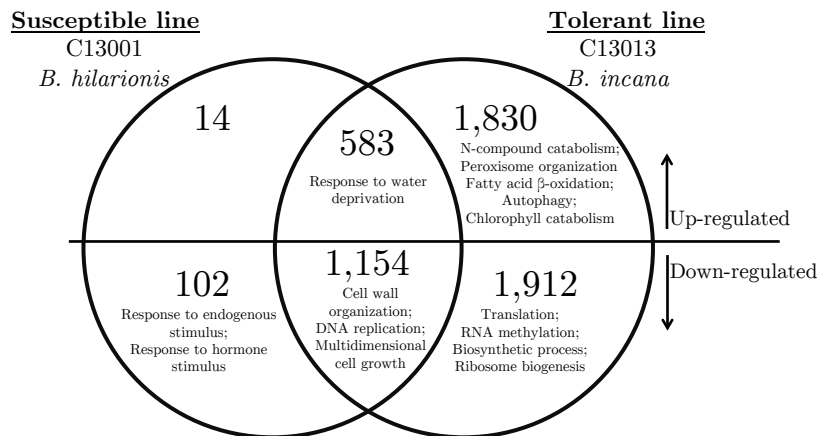


Figure 5.9: Venn diagram to illustrate the overlap of up-/down-regulated differentially expressed in response to salt shock in a susceptible and tolerant line of wild C-genome *Brassica* species

Overlap of up- and down-regulated differentially expressed genes in lines C13001 (*B. hilarionis*; susceptible) and C13013 (*B. incana*; tolerant). The number of genes in each category is shown along with the top over-represented GO terms (Biological Process category only; $p < 0.05$) for each category as determined using BiNGO (Maere et al., 2005b).

in all lines of *Brassica* under salt shock. Of these, 22 genes were differentially expressed in all of the *Brassica* lines suggesting they belong to a ‘core’ set of genes in response to salt shock.

The low number of differentially expressed genes seen in this line could be because the excess Na^+ ions are prevented from reaching the shoot. Therefore, investigating the root transcriptome of this line in response to salt shock, may uncover interesting mechanisms of salinity tolerance in *Brassica*.

Due to the small number of differentially expressed genes in this line, C10128 was not included in subsequent analyses.

5.2.9 Analysing genes differentially expressed in susceptible and tolerant lines

The overlap between numbers of genes differentially expressed in the susceptible (C13001) and tolerant (C13013) lines can be seen in the Venn diagram, for both up- and down-regulated genes (Fig. 5.9).

There was a common set of 1,737 genes (583 up-regulated and 1,154 down-regulated) which were differentially expressed in both lines, representing a core response to salt shock regardless of tolerance level. The up-regulated core genes were associated with the response to water deprivation whilst the down-regulated genes were associated with a wider range of biological processes, including cell wall organisation, DNA replication and

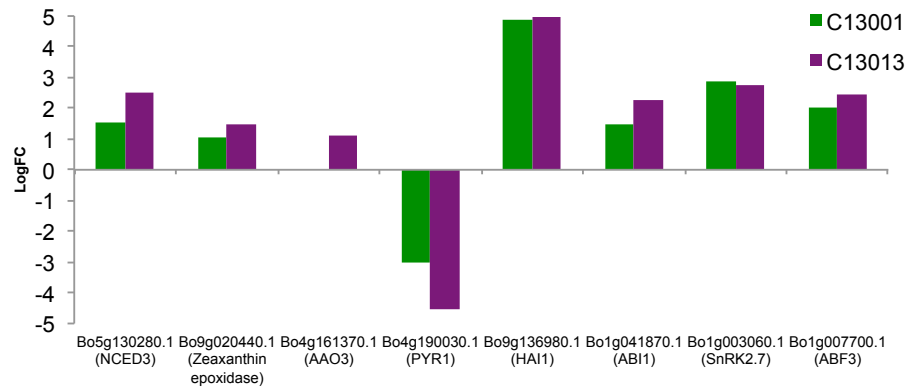
cell proliferation. This group of genes may be responsible for the general response to salt shock that was discussed previously in GD33DH, such as a decrease in growth (Chapter 3), disruption to the cell wall and a decrease in transcription (Fig. 4.9).

A small number of 116 genes were differently expressed only in the susceptible line C13001, the majority of which were down-regulated. This group of genes was associated with a response to endogenous stimulus and hormone stimulus suggesting general roles in stress signalling and response.

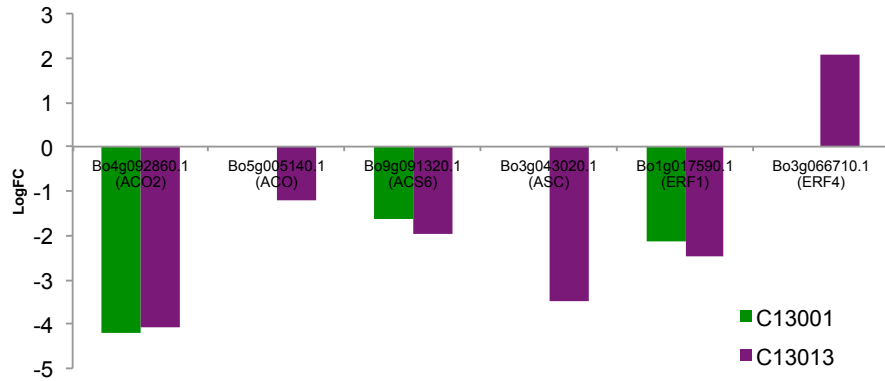
Due to the larger number of differentially expressed genes in the tolerant line (C13013; Table 5.3), there were substantially more genes differentially expressed only in C13013. 1,830 of these were up-regulated and 1,912 were down-regulated, totalling 3,742 genes which may be responsible for conferring the enhanced tolerance seen in the C13013 line (Fig. 5.6 and Fig. 5.9). There were numerous varied GO terms associated with the genes differentially expressed only in C13013. The up-regulated genes were associated with biological processes such as nitrogen compound catabolism, fatty acid beta-oxidation and chlorophyll catabolism, and GO terms associated with down-regulated genes such as translation, RNA methylation, biosynthetic process and ribosome biogenesis (Fig. 5.9) indicates that a large alteration in metabolism may be occurring at the transcriptional level as excess Na^+ ions affect metabolism. In addition, the effects of salt shock are evident in the up-regulation of genes associated with autophagy suggest that the plant induces programmed cell death where levels of Na^+ exceed a threshold of toxicity.

The biological processes associated with the differentially expressed genes showed similarities to those seen in GD33DH following salt shock suggesting that some of the core and salt tolerance characteristics are present in GD33DH (Fig. 4.9). As was seen in the intersection of C13001 and C13013 differentially expressed genes, cell wall organization and multidimensional cell growth were also found to be down-regulated in GD33DH between 2-4hpt. When comparing GO terms associated with the genes differentially expressed in C13013 with GD33DH expression, up-regulation of autophagy and peroxisome organisation was initiated between 14-18 hpt. In addition, down-regulation of photosynthesis, translation, ribosome biogenesis and biosynthetic processes occurred in GD33DH following salt shock (Fig. 4.9).

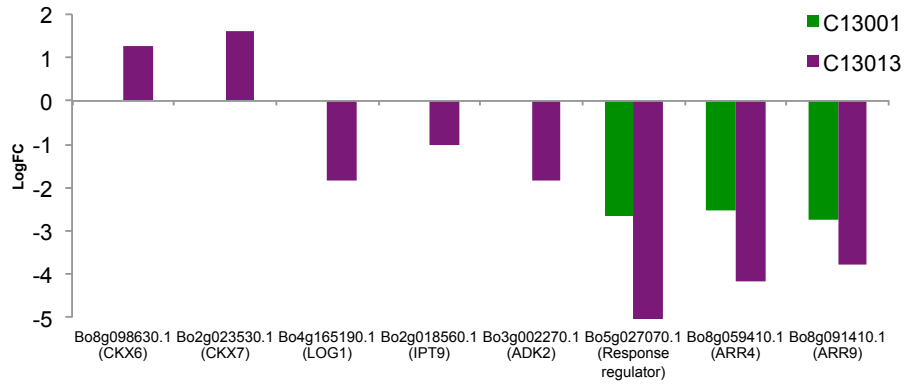
This analysis suggests that transcriptional reprogramming and the early down-regulation of metabolic and biosynthetic processes are mechanisms favoured by the tolerant line C13013 and may result in less severe growth retardation in the later stages following salt shock (Fig. 5.6). In the section that follows, groups of genes relating to specific biological processes were compared between tolerant and susceptible lines in an attempt to identify the mechanisms conferring enhanced tolerance to the C13013 line.



(a) ABA



(b) Ethylene



(c) Cytokinin

Figure 5.10: Expression of a selection of hormone related genes in salt shock susceptible and tolerant *Brassica* species

Green represents the susceptible line C13001 (*B. hilarionis*), purple represents the tolerant line C13013 (*B. incana*). Individual genes are on the *x*-axis, logFC on the *y*-axis. Hormone related genes expression includes (a) ABA, (b) ethylene and (c) cytokinin.

Hormone signalling pathways

Plant hormones such as ABA, ethylene and cytokinins are essential for plant adaptation to abiotic stress conditions, as shown previously (Fig. 4.12). In both susceptible and tolerant lines, many genes that may be involved in hormone biosynthesis and signalling showed altered expression in response to salt shock (Fig. 5.10). Genes encoding orthologs of key enzymes involved in the biosynthesis of ABA (*NCED3*, Zeaxanthin epoxidase and *AAO3*) (Barrero et al., 2006) generally showed consistent patterns of expression between both C13001 and C13013, with the tolerant line C13013 showing slightly higher logFC compared to C13001, except in the instance of *AAO3*, which was only differentially expressed in C13013 (Fig. 5.10a). Expression of these genes in the wild *Brassica* lines was comparable with the expression patterns shown in GD33DH (Fig. 4.10). Differentially expressed genes encoding key components of the perception of ABA such as *PYR1*, *HAI1*, *HAI2* (data not shown), *ABI1*, *ABI2* (data not shown), *SnRK2.7*, *ABF3* and *ABF4* (data not shown) (Bhaskara et al., 2012; Chan, 2012; Leung et al., 2002; Yoshida et al., 2010). This showed consistent expression patterns between C13001 and C13013 (Fig. 5.10a). These results suggest that ABA signalling under salt shock is consistent between species and may not be the cause of the additional tolerance seen by C13013.

Interestingly, the majority of ethylene related genes were down-regulated following salt shock (Fig. 5.10b). This included genes involved in ethylene biosynthesis such as *AC02* and *ACS6* which were down-regulated in both C13001 and C13013. In addition, *ASC* and *ACO* were down-regulated in only C13013 and downstream ethylene responsive TFs such as *ERF1* was down-regulated in both lines. *ERF4*, a transcriptional repressor capable of modulating the ABA and ethylene (Yang et al., 2005) was up-regulated in C13013. This suggests that ABA and ethylene work antagonistically in response to salt shock in the susceptible and tolerant lines, and that ABA signalling is favoured over the ethylene signalling pathway at 24 hpt.

A collection of Type A response regulator proteins including ones with orthology to *ARR4* (Fig. 5.10c) proteins were highly down-regulated in both susceptible and tolerant lines with stronger down-regulation in C13013. A reduction of cytokinin levels, together with ABA regulation of stomatal aperture may have a role in the adaption to drought stress (O'Brian and Benková, 2013). Overexpression of *ARR4* results in increased shoot growth (Osakabe et al., 2002) and *ARR9* has been shown to be regulated by the circadian clock in a cytokinin independent manner (Ishida et al., 2008). The Arabidopsis *ARR1*, *ARR10* and *ARR12* proteins have recently been implicated in the negative regulation of the cytokinin signalling pathway in response to drought (Nguyen et al., 2016b).

In addition, genes with orthologs to cytokinin oxidase (*CKX6*, *CKX7*), enzymes that are involved in the degradation of cytokinin were found to be up-regulated in the tolerant C13013 line. Salt induced senescence has been associated with low cytokinin and high

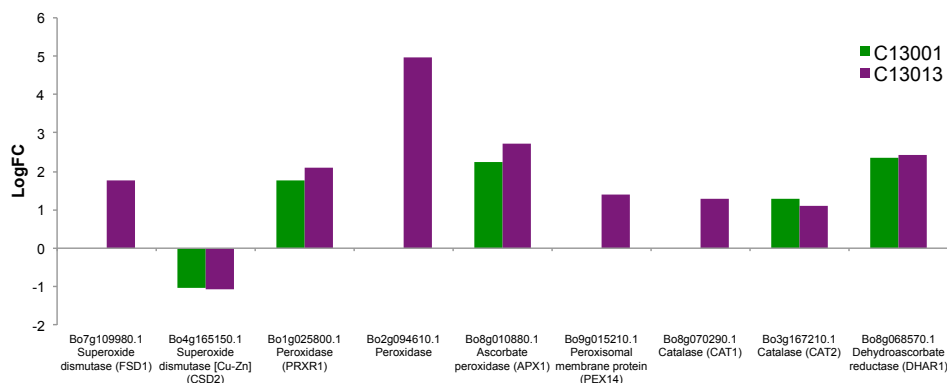


Figure 5.11: Expression of a selection of differentially expressed antioxidants genes in salt shock susceptible and tolerant *Brassica* species

Green represents the susceptible line C13001 (*B. hilarionis*), purple represents the tolerant line C13013 (*B. incana*). Individual genes are on the *x*-axis, logFC on the *y*-axis.

ABA levels (Ghanem et al., 2008). The decreased expression of cytokinin biosynthesis genes suggests that cytokinin levels are reduced in leaves responding to salt shock in *Brassica* which may result in decreased shoot growth and early onset senescence in older leaves.

That multiple hormone biosynthesis and signalling pathways are differentially regulated in the salt susceptible and tolerant lines suggests complex cross talk between each pathway, resulting in the fine tuning of downstream genes that are important in orchestrating tolerance to salt shock in *Brassica* species.

Antioxidants

Under stress conditions, protection against further oxidative stress caused by excess ROS is important in protecting the cells from damage leading to programmed cell death. Antioxidants such as SOD, peroxidases and catalases are enzymes that scavenges excess ROS, which is reduced to O₂ to protect the cell from the effects of oxidative stress. Several genes encoding antioxidants were differentially expressed in both lines, including genes with orthology to *CSD5*, *PRXR1*, *APX1*, *CAT2* and *DHAR1*. However, there were more differentially expressed antioxidant genes in the tolerant C13013 line for example, genes with orthology to *FSD1*, *PEX14* and *CAT1* (Fig. 5.11).

Superoxide dismutase genes localised within the chloroplast, such as *FSD1* (Fe²⁺ SOD) were up-regulated only in C13013 and others such as *CSD2* (Cu²⁺/Zn²⁺ SOD) were down-regulated in both lines. It has been previously shown that expression of *FSD1* is differentially expressed in a genotype specific manner in different accessions of *Arabidopsis* under salt stress which was linked to tolerance (Attia et al., 2011). Reduced levels of *CSD2*, as regulated by miR398 (Sunkar et al., 2006) has been shown to be implicated

in thermotolerance (Lu et al., 2014) and mild salt stress (Attia et al., 2008), both in *Arabidopsis*.

An increase in peroxisome proliferation is seen in plants under salt stress (Hernandez et al., 1995), however no link has yet been established between peroxisome proliferation and salt tolerance in *Arabidopsis* (Mitsuya et al., 2010). The peroxisomal ascorbate peroxidase gene *APX1* was up-regulated in both lines following salt shock. The APX gene was cloned from the extreme halophyte *Salicornia brachiata* (*SbpAPX*) and confers salt tolerance and protection against oxidative stress in both transgenic tobacco and peanut (Singh et al., 2013, 2014). This may suggest that lipid oxidation enzymes are important for ROS signalling and the increased presence of peroxisome-related gene expression to scavenge the excess ROS may be responsible for the enhanced tolerance seen in C13013. Indeed, in this line two ascorbate peroxidases and three peroxidases were found up-regulated in response to salt shock, suggesting potential mechanisms for increased tolerance in C13013.

Catalases (encoded by *CAT1* and *CAT2*) are able to function without the need of a reductant and have key roles in H_2O_2 metabolism (reviewed in Mhamdi et al., 2010). *CAT1* (Class III catalase) was up-regulated only in C13013 and *CAT2* was up-regulated in both lines, suggesting that the additional up-regulation of *CAT1* in C13013 may contribute to the enhanced tolerance in C13013. Dehydroascorbate reductase (DHAR1) catalyses the regeneration of ascorbate, an important antioxidant, from its oxidised state. This has implications for tolerance to oxidative stress, growth and development of leaves (Chen and Gallie, 2006). The gene was up-regulated in both lines suggesting it is important in the core response to salt shock in *Brassica* species.

Increased expression of a greater number of antioxidant genes in the tolerant line C13013 may be responsible for an enhanced ability to scavenge ROS, resulting in the enhanced tolerance to salt shock seen in this line, through a reduction of oxidative stress and damage caused by excess ROS.

Photosynthesis and respiration

The photosynthetic response to high salinity is complex and is affected in many ways, for instance through decreased CO_2 availability caused by stomatal closure and through alterations of photosynthetic metabolism by damage to the photosynthetic machinery caused by oxidative stress (reviewed in Chaves et al., 2008). Genes associated with photosynthesis were found to be down-regulated in GD33DH 18-24 hpt (Fig. 4.9).

Multiple genes encoding chlorophyll A/B binding proteins associated with the photosystem II (PSII) complex were down-regulated under salt shock, as is seen in response to drought and salinity in *Arabidopsis* (Chaves et al., 2008; Kilian et al., 2007). The tolerant line C13013 had a greater number of down-regulated PSII genes compared to the susceptible line C13001 suggesting a greater response regarding PSII in C13013. Repair of PSII

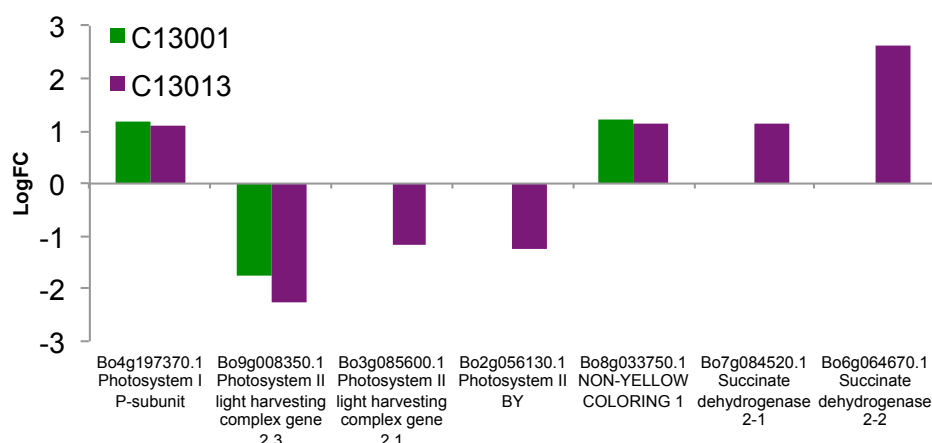


Figure 5.12: Expression of a selection of differentially expressed genes involved in photosynthesis and respiration in salt shock susceptible and tolerant *Brassica* species

Green represents the susceptible line C13001 (*B. hilarionis*), purple represents the tolerant line C13013 (*B. incana*). Individual genes are on the *x*-axis, logFC on the *y*-axis.

following photoinhibition is highly sensitive to unfavourable conditions such as high salinity, and inhibition of protein synthesis of PSII subunits is seen in *Synechocystis* cyanobacteria under salt stress conditions (Allakhverdiev and Murata, 2004; Allakhverdiev et al., 2002). Contrary to previous observations in GD33DH, a gene encoding a Photosystem I subunit was up-regulated in both lines following salt shock (Fig. 5.12).

In addition to the effects of high salinity on PSI and PSII, a gene encoding a protein involved in the degradation of chlorophyll (*NYE1*) (Ren et al., 2007) was up-regulated in both lines under salt shock. The differential expression of such genes indicates down-regulation of photosynthesis, particularly in the tolerant line which could act as a key mechanism of tolerance under high salt conditions in *Brassica*.

The succinate dehydrogenase genes (*SDH2-1* and *SDH2-2*) were up-regulated in the tolerant line C13013. These proteins are involved in respiration and belong to complex II located in the mitochondria. SDH has roles in the generation of ROS for downstream signalling, regulating both development and stress response in *Arabidopsis* and *O. sativa* (Jardim-Messeder et al., 2015). This suggests that whilst photosynthesis is affected by salt shock through down-regulation of Photosystem II, respiration is up-regulated in the tolerant line C13013 possibly providing additional energy that is required to balance the ionic content of the cell.

Ion transporters

Another group of genes that showed altered expression in response to salt stress encode proteins relating to transport and cellular homeostasis (Fig. 5.13). Salt stress not only imbalances Na^+ and Cl^- ions, it also has an effect on the homeostasis of other intracellular

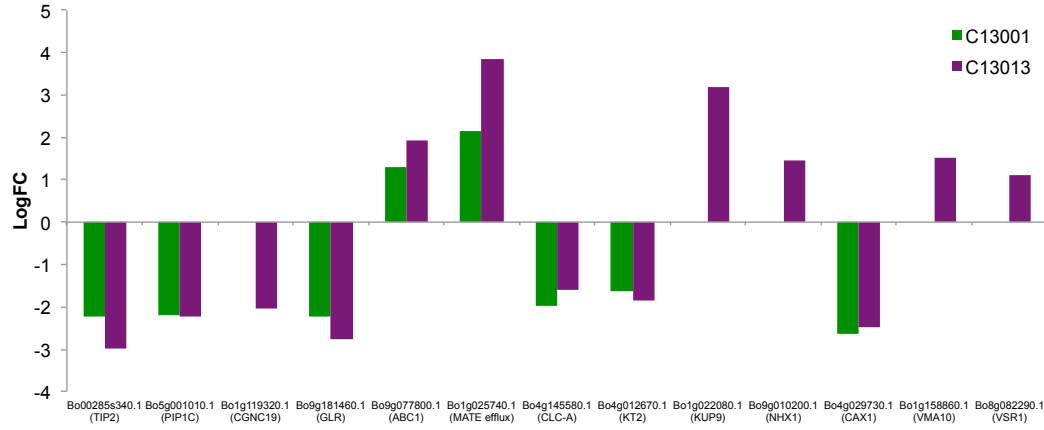


Figure 5.13: Selection of differentially expressed ion transporter genes in salt shock susceptible and tolerant *Brassica* species

Green represents the susceptible line C13001 (*B. hilarionis*), purple represents the tolerant line C13013 (*B. incana*). Individual genes are on the *x*-axis, logFC on the *y*-axis.

ions such as Ca^{2+} , K^{+} , and NO_3^{-} . Responding to the need to rebalance the intracellular ionic content to prevent cytotoxic effects following salt shock, differential expression of various transporter proteins was seen in the *Brassica* lines.

There was an abundance of both up- and down-regulated ion transporters in the *Brassica* species under salt shock conditions. This included a large collection of non-selective cation channels (NSCCs) such as ABC transporter proteins, aquaporins (TIPs and PIPs), CNGCs and glutamate receptors were down-regulated under salt shock a selection of which are shown in Figure 5.13. NSCCs allow the passive transport of Na^{+} and other ions past the plasma membrane or tonoplast (Maathuis, 2013). Down-regulation of these genes may be a potential mechanisms that plants use to prevent uptake of excess Na^{+} ions into the cytoplasm.

MATE efflux family and ABC transporter proteins were up-regulated in both lines, suggesting an important role for these transporters in the regulation of ion homeostasis. The maintenance of a high $\text{K}^{+}:\text{Na}^{+}$ ratio is important in tolerance to salt shock. Many potassium transporters such as those encoding orthologs to *KUP9*, *KUP11*, *KAB1* and *KCO5* were up-regulated, whilst some such as *KT2* were down-regulated.

Sequestration of ions within the vacuole is clearly an important mechanism of salt tolerance in C13013, with many vacuolar transporters showing differential expression. A $\text{Na}^{+}/\text{H}^{+}$ exchanger gene with orthology to the *NHX1* antiporter was up-regulated in C13013 only. The *NHX1* gene is essential for K^{+} homeostasis in Arabidopsis, actively transporting K^{+} into the vacuole to maintain turgor pressure (Bassil et al., 2011). In a *nhx1 nhx2* double mutant exposed to salinity stress, sodium accumulated to significantly greater levels in the mutant compared to control, suggesting the *NHX1* and *NHX2* genes (which are functionally redundant) also play a role in Na^{+} compartmentalisation (Barragan

et al., 2012). Interestingly, a vacuolar cation calcium exchanger with orthology to *CAX1* was down-regulated in both lines. A *cax1* mutant shows an impairment of ion homeostasis, altered development and response to hormones (Cheng et al., 2003, 2005), indicating an important role for this transporter. A selection of vacuolar ion transporters were differentially expressed only in the tolerant C13013 line, including a V-type proton ATPase (*VMA10*) and Vacuolar-sorting receptor 1 (*VSR1*) suggesting sequestration of intracellular Na^+ ions within the vacuole, protecting cells from cytotoxicity caused by high Na^+ content.

The differential expression of transporters suggests complex interplay between transporter proteins in order to maintain cellular homeostasis. There were few differences between lines for NSCCs, MATE efflux and ABC transporter proteins suggesting that prevention of ion uptake and ion detoxification is a core mechanism in the defence against salt shock. However, the abundance of potassium transporters in the tolerant line C13013 may be responsible for maintaining the high K^+ , low Na^+ ratio within the cytosol, which is essential for salt tolerance (Fig. 5.7).

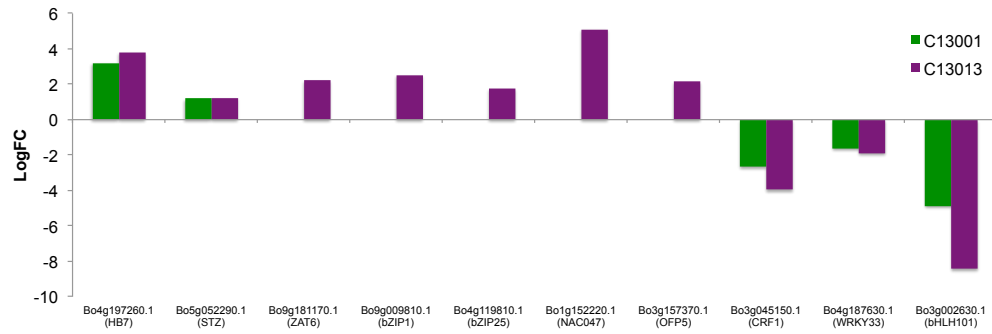
5.2.10 Transcription factors regulating salt shock response have genotype specific expression patterns

TFs play an important role in modulating the response of plants to severe environmental conditions such as high salinity. Analysis of the differentially expressed genes revealed a large collection of TFs potentially responsible for orchestrating a response to salt shock (Table 5.3). The expression patterns of TFs were interesting, as highlighted in Figure 5.14. Many TFs were differentially expressed as a single copy i.e. one copy of each Arabidopsis ortholog, particularly in the susceptible line C13001 (Fig. 5.14a and b). However, in the tolerant line C13013, it was the case that many TFs were differentially expressed in multiple copies with two or more genes per Arabidopsis ortholog (Fig. 5.14b) suggesting that the increased tolerance seen in C13013 could be as a result of tighter modulation of gene expression as a result of having multiple copies of TFs expressed.

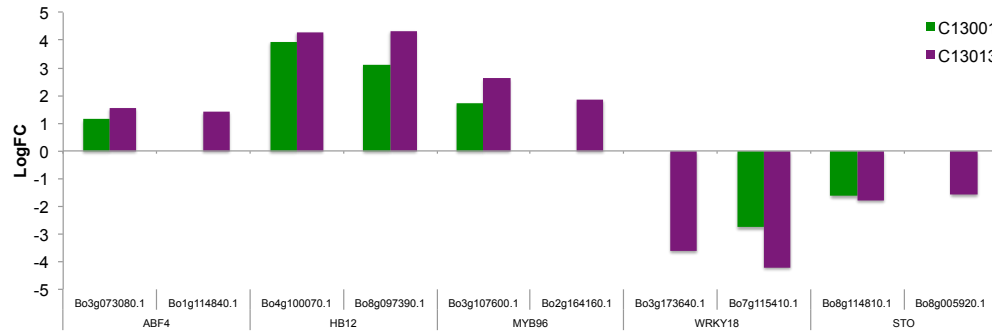
Over-expression of the *STZ* (or *ZAT10*) induces the expression of several salt stress genes and also induces growth retardation and tolerance to drought stress (Mittler et al., 2006; Sakamoto et al., 2004). This gene is up-regulated in both lines of *Brassica* suggesting a core role in growth retardation following salt shock.

bZIP1 and *bZIP25* were up-regulated only in the tolerant C13013 line. *bZIP1* alongside *bZIP53* and *bZIP25* plays an established role in the response to salinity stress in Arabidopsis, reprogramming primary C and N metabolism (Hartmann et al., 2015).

The ABF3 (Fig. 5.10) and ABF4 (Fig. 5.14) proteins are part of a core of master transcription factors that regulate ARBE-dependent ABA signalling under water stress conditions (Yoshida et al., 2010), with down-stream targets such as PP2Cs, LEA class



(a) TFs expressed in singular copies



(b) TFs expressed in multiple copies

Figure 5.14: Selection of differentially expressed TFs in salt shock susceptible and tolerant *Brassica* species

Green represents the susceptible line C13001 (*B. hilarionis*), purple represents the tolerant line C13013 (*B. incana*). Individual genes are on the *x*-axis, logFC on the *y*-axis. (a) TFs differentially expressed as a single gene copy i.e. one Arabidopsis ortholog; (b) TFs expressed in multiple copies i.e. two or Bo genes per Arabidopsis ortholog.

genes and many TFs, as discussed previously (Chapter 4). Both genes were up-regulated in both susceptible and tolerant lines, suggesting, as described previously the importance of ABA in the core response to salt stress.

In *Arabidopsis*, the HB7 and HB12 proteins have been predicted to have evolved divergently to fine tune processes associated with growth and the response to water stresses (Ré et al., 2014) by repressing growth and regulating the exclusion of sodium ions from the cell cytosol suggesting the importance of these genes in the core response to salt shock. HB12 has been previously associated with enhancing salinity tolerance in yeast by regulating sodium exclusion (Shin et al., 2004) and has also been associated with the negative regulation of the growth of the inflorescence stem (Son et al., 2010). The *HB7* (Fig. 5.14a) and *HB12* (Fig. 5.14b) genes were up-regulated TFs in both lines and therefore must play a key role in the response to salt shock.

A gene with orthology to the cytokinin response factor (*CRF1*) (Fig. 5.14a) is also down-regulated in both lines, suggesting that cytokinin biosynthesis and signalling is suppressed in the core response to salt shock. bHLH101 has roles in the regulation of iron homeostasis in *Arabidopsis* (Sivitz et al., 2012) and has found to be down-regulated under salt stress conditions in *B. juncea* under salt stress (Sharma et al., 2015). A gene with orthology to *bHLH101* was highly down-regulated in both lines under salt shock, suggesting an important role for this gene (Fig. 5.14a). WRKY18 is a weak transcriptional activator that forms a complex with WRKY40 and WRKY60 to negatively regulate the response to ABA in abiotic stresses and the defence response (Chen et al., 2010; Shang et al., 2012). *WRKY18* is down-regulated in both lines, in two copies in the C13013 line (Fig. 5.14b), suggesting that the down-regulation of this gene important in the activation of the ABA signalling pathway in *Brassica* under salt shock conditions.

5.2.11 Over-represented TF families in salt shock susceptible and tolerant *Brassica* lines

To gain a more general picture of TF expression in response to salt shock in the susceptible and tolerant *Brassica* lines, TF families potentially responsible for controlling the complex network of signalling pathways in response to salt shock were investigated. TF families were analysed for over-representation when compared to background TO1000 genome using a hyper-geometric test shown in Figure 5.15. Although the usual collection of stress related TFs were present in abundance (AP2-EREBP, bHLH, HB, MYB and WRKY), they were not found to be over-represented in wild *Brassica* in response to salt shock above background. Members of these TF families have been previously reported to have a role in the abiotic stress response (Liu et al., 2015a; Peng et al., 2014) and were discussed in Chapter 4.

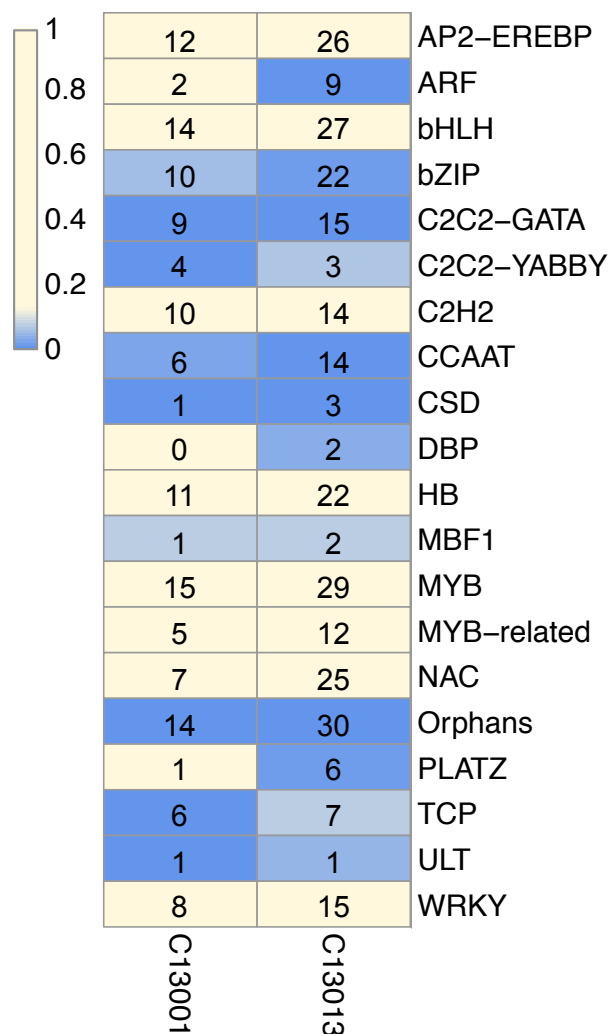


Figure 5.15: Enrichment analysis of selected TF families in susceptible and tolerant *Brassica* lines following salt shock treatment

TF families were identified based on homology to the Plant Transcription Factor Database (PlnTFDB 3.0; Jin et al., 2013; Perez-Rodriguez et al., 2009). Over-representation analysis was carried out using a hypergeometric test as described in Chapter 2. The resulting p -value is indicated as per the scale. The number within each box represents the number of differentially expressed TF family members per line.

Smaller, less well characterised TF families such as CSD, Orphans, PLATZ, CCAAT and TCPs were found to be over-represented in *Brassica* (Fig. 5.15). For instance, the CSD (Cold Shock Domain) TF family have high homology to bacterial cold shock proteins and function as transcription antiterminators or translational enhancers by destabilizing RNA secondary structures (Nakaminami et al., 2006). *Arabidopsis* has four genes in this family, which are differentially regulated in response to low temperature. There were 3 *Brassica* CSD TFs that were significantly up-regulated in C13013, whilst only one gene from this family was up-regulated in C13001 suggesting that these genes may play a role in tolerance to salt shock, in addition to cold stress.

AUXIN RESPONSE FACTORS (ARF) regulate auxin responses by binding to auxin response DNA elements (AuxRE) in the promoters of auxin-regulated genes and either activate or repress transcription (reviewed in Li et al., 2016) in response to biotic and abiotic stressors (Wang et al., 2010). In *Camellia sinensis*, several ARF genes were up-regulated under salt stress conditions in the shoots (Xu et al., 2016). ARFs are over-represented in the tolerant C13013 line, but not in the susceptible C13001 line suggesting gene expression regulation by auxin is important in C13013 and may be partially accountable for the differential growth response seen in the *Brassica* lines in the phenotype screen (Fig. 5.6).

The CCAAT TF family (consisting of the NF-Y genes) contains up to 14 differentially expressed TFs in the *Brassica* lines. The proteins encoded by these genes assemble into a heterotrimeric complexes of NF-YA, NF-YB and NF-YC subunits which bind to the CCAAT motif and regulate expression of downstream genes. These genes have been found differentially expressed in various stress conditions, including endoplasmic reticulum stress (Liu and Howell, 2010), senescence (Breeze et al., 2011) and in response to infection with *B. cinera* (Windram et al., 2012). The majority of CCAAT TF family genes were mildly up-regulated only in C13013. Within the CCAAT family of TFs, there were two Histone 2A proteins that were down-regulated in both lines. Under abiotic stress conditions, chromatin remodelling is an effective method of transcriptional regulation through nucleosome disassembly and the down-regulation of Histone 2A, may indicate evidence of this process in both lines of *Brassica*.

The Orphans group of TFs consists of 117 genes in the *B. oleracea* TO1000 genome. This is a collection of TFs which did not align readily to a TF family, however were well characterised and so were placed into an ‘Orphans’ family. In this experiment, the differentially expressed TFs contained within the Orphans group mainly consisted of down-regulated B-box zinc finger family proteins, down-regulated response regulators (as discussed above) and up-regulated zinc finger proteins (CONSTANS-like protein). A large number of the zinc finger proteins were differentially expressed only in C13013 suggesting that they are important in salt shock tolerance.

Little is known about the PLATZ TF family, PLATZ1 was originally isolated from

peas (Nagano et al., 2001). The *GmPLATZ1* gene was found to be up-regulated by abiotic stresses such as drought, salt or ABA applications in soybean (*Glycine max* L.) (So et al., 2015) and a PLATZ TF was found to be up-regulated in the shoot of *Sorghum bicolor* in response to osmotic stress (Dugas et al., 2011). Six highly up-regulated members of the PLATZ TF family are over-represented in C13013 and therefore are likely to play a key role in the response to salinity/osmotic stress, though more research would be necessary to establish a role.

Finally, the TCP family is a large plant-specific family of developmental regulators with roles in cell proliferation, growth (Li et al., 2005) and biotic stress defence (Windram et al., 2012). Between 6 and 7 genes per line were down-regulated during the salt response in both *Brassica* lines, suggesting an important role for this group of genes relating to the decreased growth and development seen in the salt-treated plants (Fig. 5.6).

5.2.12 Comparing gene expression of susceptible and tolerant lines to the DHSL150 rapid cycling parent

In order to compare differential gene expression to the DHSL150 parent line, the expression profiles of the 5,582 genes that were differentially expressed in either C13001 or C13013 under salt shock (Table 5.9) were analysed and a heat map of logFC expression was produced using hierarchical clustering. A cut was made along the dendrogram to cluster the genes into 6 clusters, representing distinct patterns of expression between the susceptible and tolerant lines compared to the DHSL150 parent (Figure 5.16).

Cluster A (1,302 genes) was the largest cluster and showed general down-regulation of genes in both C13001 and C13013, with a general up-regulation of genes in the parent DHSL150 line. Cluster B (1,136 genes) showed a considerable down-regulation of gene expression in C13013, but only moderate down-regulation of gene expression in C13001 and DHSL150. Cluster C (467 genes) showed no changes in C13001 and slight down-regulation of expression in C13013 and DHSL150. Of the up-regulated clusters, Cluster D (707 genes) showed strong up-regulation of genes in both C13001 and C13013 with a lower induction of expression in DHSL150. Cluster E (1,083 genes) showed a high induction of expression in the tolerant line C13013 with weaker induction of expression in both C13001 and DHSL150. Finally Cluster F (887 genes) showed a strong induction of expression in genes from both C13013 and DHSL150, with weaker up-regulation in C13001 (Figure 5.16).

The most striking observation from this figure is the impact of the carrying out crosses between the wild species x DHSL150 on the epigenetic regulation of gene expression. Cluster A shows the largest difference in differential gene expression between the DH lines (both showing down-regulated expression) and the DHSL150 parent line (up-regulated expression). There was little difference between the two wild lines C13001 and C13013.

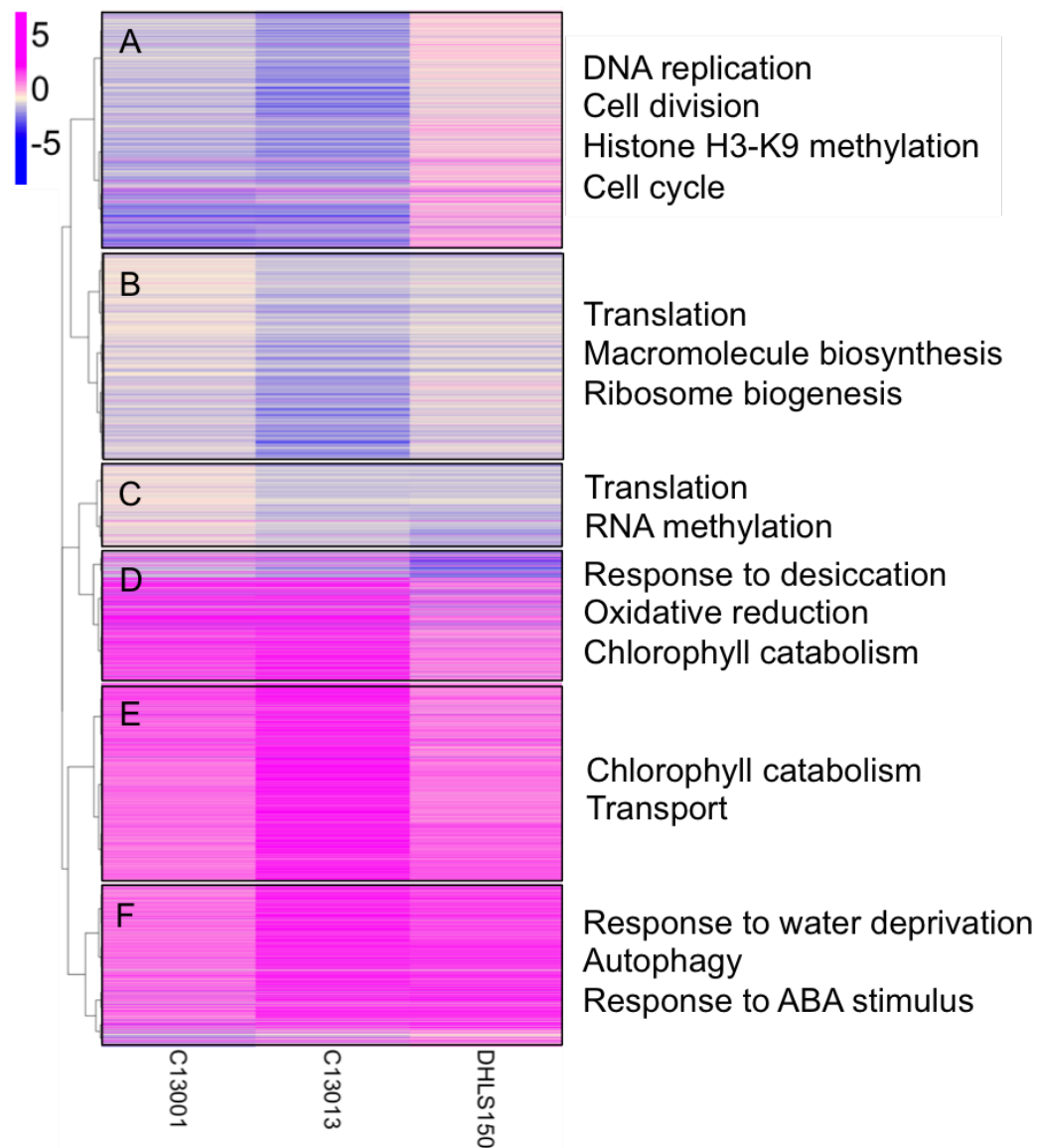


Figure 5.16: Complete linkage hierarchical cluster analysis of salt shock induced changes of gene expression in three lines of *Brassica*

Heat map shows the logFC in gene expression between treatment and control of three lines of *Brassica* - C13013 (*B. incana*), C13001 (*B. hilarionis*) and DHSL150 (*B. oleracea*). Scale is logFC ranging from pink (highly up-regulated) to blue (highly down-regulated). Six clusters were formed (labelled A-F, highlighted in black) and over-represented GO terms for each cluster, if any, are indicated at the bottom of each cluster.

Over-represented GO terms associated with genes in this cluster included DNA replication and Histone H3-K9 methylation, suggesting an effect on epigenetic regulation. By contrast, other clusters showed that in C13001 and C13013, expression of genes in Clusters B-F is similar to that of the DHSL150 parent, at least in direction of expression.

5.3 Discussion

Cultivated crop species often lack the genetic plasticity to cope with extreme conditions, following years of selective breeding in which traits which may be useful in adaption have been lost. Wild species generally show greater variation in response to stress conditions, thus their germplasm may be useful in widening the gene pool of cultivated species to improve their tolerance to abiotic stress conditions (Henry, 2014). The availability of diverse germplasm such as the Cg-DFS and the Cg-DFFS provide an excellent opportunity to understand the genetic variability of wild C-genome *Brassica* species in response to abiotic stress. This collection has been previously screened for variation in multiple traits such as shoot Ca^{2+} and Mg^{2+} concentration (Broadley et al., 2008; White et al., 2010), seed oil content (Barker et al., 2007) and water use efficiency (Thompson, 2011, 2009) and was found to exhibit excellent genetic variability.

The analysis of physiological characteristics of a range of wild C-genome *Brassica* species (S1 produced from selfed founder lines, and DH lines produced by crossing the founder lines and the rapid cyeler DHSL150), showed that the wild *Brassica* species altered their physical characteristics in varying ways as a result of salt shock. Some lines were highly susceptible to salt shock, exhibiting a significant reduction in both height and fresh/dry weight compared to controls. Others showed a varied response with reduced height, but a maintenance of fresh/dry weight or leaf area.

Growth inhibition may be explained by the reduced photosynthetic capabilities of plants under salt stress due to the closure of the stomata, and limited water uptake capacity due to a high osmotic potential outside of the root system. In addition, the damage caused within the shoots by excess Na^+ ions causes an osmotic stress which in some cases results in cell death and salt-induced senescence. The increase in whole plant dry weight could be as a result of increased solute content of the leaf, or increased thickness of the leaf cause by an increase in leaf surface wax. There is conflicting evidence on the effects of salt stress on plant biomass, with some sources reporting an increase in biomass (Memon et al., 2010; Qados, 2011), whilst others show a decrease (Sairam et al., 2002; Siddiqui et al., 2015, 2008) dependent on species and severity of stress.

Understanding the molecular basis behind salt tolerance in contrasting genotypes may help in the development of stress tolerant varieties of *B. oleracea*. A global transcriptional reprogramming is often seen in plants responding to stress conditions (Buchanan-Wollaston

et al., 2005; Liu et al., 2015b; Peng et al., 2014; Windram et al., 2012). Following extensive phenotyping of wild C-genome *Brassica* S1 founder lines, and DH lines, three DH lines representing wild C-genome *Brassica* species, C13001 (*B. hilarionis*; susceptible), C10128 (*B. oleracea*; tolerant) and C13013 (*B. incana*; tolerant) were selected for sequencing, alongside the DHSL150 (*B. oleracea*) which was used as a parent in the generation of the DH lines. The abundance of many genes associated with multiple biological pathways and functions was found to be altered in wild Cg *Brassica* species, revealing a complex transcriptional network controlling the response to salt shock.

The results gained from the C10128 analysis were interesting. Under salt shock conditions, this line performed exceptionally well in terms of retaining fresh/dry weight following salt shock, although some height and leaf area was lost following treatment (Fig. 5.6). The number of differentially expressed genes was few (Table 5.4), however the majority of genes that were differentially expressed in this line were also differentially expressed at high levels in at least one other line of *Brassica* following salt shock. When this is combined with the high level of tolerance seen in this line (Fig. 5.6) and the high $K^+ : Na^+$ ratio (Fig. 5.7), it is possible that C10128 is highly efficient in dealing with the effects of salt shock such that by 24 hpt, the majority of gene expression alterations have already taken place, a high $K^+ : Na^+$ ratio is established and growth is relatively unaffected. It is possible that these genes are highly important in salt shock tolerance and should be further investigated as potential breeding targets for the development of tolerant lines of *Brassica*.

The tolerant C13013 had a greater number of differentially expressed genes compared to the susceptible C13001 line, possibly suggesting a more diverse transcriptional response of C13013, conferring enhanced tolerance to salt shock. Comparison of gene expression between susceptible and tolerant lines is important in determining a core set of salt responsive genes and genes that are expressed only in the tolerant line and may be more likely to be the cause of enhanced tolerance. In addition, the logFC was generally higher showing stronger induction or reduction in gene expression in C13013 compared to C13001. C13013 also showed expression of many more important salt stress related genes (e.g. Fig. 5.10, Fig. 5.11 and Fig. 5.13), a great alteration in metabolism (Fig. 5.12) and the differential expression of a wider range of TFs which are capable of tightly modulating the down-stream response to salt shock (Fig. 5.14 and Fig. 5.15). A similar study was carried out in barley under salinity stress in which contrasting genotypes were analysed. Distinct groups of genes were expressed in each of the lines, and were considered either as salt responsive or in contributing to salinity tolerance (Gao et al., 2013; Guo et al., 2009).

The analysis revealed a collection of 1,737 core response genes which were differentially expressed in both tolerant and susceptible lines with biological functions relating to response to water deprivation, cell wall organization, DNA replication and multidimensional cell

growth. A large collection of 3,742 genes were differentially expressed only in C13013 and considered to be genes conferring enhanced tolerance to C13013, were associated with a wide range of GO term.

Genes implicated in hormone signalling, primarily ABA signalling, ethylene and cytokinin but also auxin, brassinosteroid, GA, SA and JA related genes were found to be differentially expressed in the wild *Brassica* species responding to salt shock. The role of hormone signalling, particularly involving ABA in response to salt stress conditions has been extensively demonstrated in the literature (Peleg and Blumwald, 2011; Raghavendra et al., 2010; Verslues, 2016; Yoshida et al., 2014b) and in the previous chapter (Chapter 4).

Interestingly, a selection of ethylene related genes were down-regulated in both lines following salt shock. The literature shows a controversial relationship between ethylene and ABA, with some studies suggesting an antagonistic behaviour under stress conditions (Picarella et al., 2007; Rosado et al., 2006; Salazar et al., 2015; Sharp and LeNoble, 2002) whilst others indicate that ABA triggers ethylene biosynthesis in fruit ripening (Zhang et al., 2009). It is likely that the relationship between these hormones alters depending on time post treatment at which sampling occurred and severity of stress as there have been reports of an increase in ethylene biosynthesis following salt acclimation in *Arabidopsis* (Shen et al., 2014).

Cytokinins have key roles in the regulation of growth and development in addition to roles in the response to abiotic stress conditions. Cytokinin levels are seen to decrease under temperature (Černý et al., 2014), drought (Kang et al., 2012) and salinity stress (Nishiyama et al., 2012). Mutants with an enhanced ability to produce cytokinin often show increased productivity under adverse conditions (Peleg et al., 2011). Genes relating to cytokinin biosynthesis (down-regulated) and cytokinin degradation (up-regulated) in the tolerant C13013 line (Fig. 5.10) may result in reduced levels of cytokinin in fully expanded leaf #5. This may result in the immediate initiation of salt induced senescence of older, fully expanded leaves within 24 hpt, possibly accounting for the decrease in fresh/dry weight seen in this line under salt shock (Fig. 5.6). This would allow the redirection of energies into the growth and development of newer leaves which are less affected by the salt shock.

Transcriptional regulation of the gene networks was under control of a range of TF families including AP2-EREBP, bHLH, bZIP, HB, MYB and NACs, many of which have been previously associated with the response to abiotic stress conditions. Interestingly, a collection of smaller, relatively unassociated transcription factor families were found to be over-represented in the *Brassica* lines, for instance, the C2C2-GATA, PLATZ and TCP families indicating crosstalk between multiple pathways resulting in complex transcriptional networks controlling downstream cellular responses to salt shock (Chan, 2012; Hartmann et al., 2015; Seki et al., 2002). It was interesting that the tolerant C13013 line had many

more TFs differentially expressed in multiple copies compared to the susceptible line. An advantage of whole genome duplication events in the evolutionary history of plants is the ability to diversify gene expression allowing for greater modulation of downstream responses to stress conditions (Roulin et al., 2012) e.g. ion transporters and antioxidants.

The impact of crossing the wild species x DHLS150 on the epigenetic regulation of gene expression is highlighted by the clustering seen in Figure 5.16. This suggests one of the major effects of crossing wild species with a rapid cycling, cultivated line is the effect on the epigenetic regulation of gene expression. Epigenetic marks are inherited, thus crossing lines with vastly different marks such as a wild with cultivated would have an effect on the epigenetic component of the resulting DH lines. It has been previously shown that chromatin changes have been detected in response to a variety of abiotic stress conditions and are important in the regulation of many stress induced alterations in gene expression (reviewed in Kim et al., 2015).

This parent-of-origin effect, or genomic imprinting has been seen previously in flowering plants (as reviewed in (Lawson et al., 2013; Rodrigues and Zilberman, 2015)). This could be the mechanism by which the long vegetative phases of the wild founder lines were altered after crossing with the rapid cycling DHSL150, rather than allelic differences, as genes with GO terms relating to cell division and cell cycle were also over-represented within this group (Cluster A, Fig. 5.16). The whole-genome effect on body mass index in offspring has been predicted in a mouse pedigree by comparing a model that incorporated parent-of-origin effects with a standard additive model (Hu et al., 2016). Although the model was not able to predict parent-of-origin effects reliably, further research and development of modelling techniques such as this could be carried out in plant populations in order to determine the effects of genomic imprinting in crop breeding.

One aspect of this experiment that was not explored in this thesis was the use of the RNAseq data to map the introgressed regions such that the parental origin of genes could be identified. This would be useful in determining which of the species was contributing the most genetic information towards tolerance against salt shock. This could be carried out through the identification of SNPs between the wild species and the DHSL150 parent line. This would constitute a useful resource for both genome-wide mapping and fine mapping of specific areas of the genome, including the possible identification of genes underlying QTL. Similar work has been carried out in *B. rapa* in which SNPs were identified in three separate lines and compared to the *B. rapa* line Chiifu (ssp. *pekinensis*) genome sequence, then a Chiifu x *B. rapa* line Tetra (ssp. *trilocularis*) RIL population was generated in which to carry out fine mapping of the *tet-o* locus (Paritosh et al., 2013).

Identifying SNPs, mapping the introgressed regions of members of the Cg-DFFS collection, and determining epigenetic effect of cross wild species with the DHLS150 rapid cyclers would be highly valuable for future researchers to determine the parental origin of

the genes which are important in the salt stress response or other research.

5.3.1 Experimental limitations and further work

The experiment does present certain limitations, that would, given more time, be resolved. These include:

- Only a single time was used in the transcriptome analysis, however it is challenging to draw conclusions on functions and regulation of genes when only one time point is considered. It would be interesting to add an additional early time point (2 hpt), this time at which a large number of transcripts are turned on in the GD33DH time-course experiment (Fig. 4.8b).
- Only four lines were sequenced. The use of more susceptible and tolerant lines in the RNAseq analysis would reveal stress tolerance mechanisms and key genes more robustly.
- Including GD33DH, 24 hpt in the RNAseq experiment would allow for more robust comparisons between the GD33DH timecourse and the wild *Brassica*.
- qPCR validation of RNAseq data to strengthen observations.
- Promoters play a key role in the regulation of gene expression and promoters found in wild species are frequently the target of investigations of plant response to stress conditions (Fischer et al., 2013). As such, a future study may focus on the role of promoters upstream of stress responsive genes in wild *Brassica* species.
- Map the points of the introgression of wild species to determine the proportion of gene expression changes comes from the wild species or from the DHSL150 parent.
- Determine the epigenetic effect of the introgression using modelling techniques and bisulphite sequencing.

5.4 Chapter Summary

A thorough analysis of the response to salt shock in a variety of C-genome *Brassica* species is presented in this chapter. A phenotype screen measuring various physiological parameters associated with the response to salt shock alongside transcriptome profiling has linked changes seen in physiology to changes in gene expression following salt shock. The results suggest that the response to salt shock is genotype dependent and tolerance mechanism vary between lines.

The results of this experiment greatly enrich the existing information on potential salt tolerant mechanisms of wild C-genome *Brassica* species and provide numerous candidate genes and suggested germplasm (C10128 and C13013) for developing salt tolerance *B. oleracea* crops.

Chapter 6

Transcriptional divergence of stress responsive paralogs in C-genome *Brassica* species

6.1 Chapter overview

Previous studies have suggested that C-genome *Brassica* species have undergone two rounds of whole genome duplication (WGD), each round followed by subsequent fractionation (gene loss) and diploidization (extensive chromosomal rearrangements) of the genome in its recent evolutionary history (Cheng et al., 2012a; Parkin et al., 2014; Wang et al., 2011b). WGD events are common occurrences in the evolutionary history of flowering plants, with estimates that 50 - 80% of angiosperms are polyploid (Wendel, 2000). By increasing the gene complement of a genome allowing for the subsequent evolution of new gene functions, polyploidy could be a mechanism for adaptation to environmental conditions and in developing tolerance to stress conditions (Vanneste et al., 2014).

The aim of this chapter is to investigate the transcriptional divergence of paralogous genes expressed in response to stress conditions in *Brassica* species. This will be achieved through the following objectives:

- Assessing differences in expression of genes originating from each sub-genome (LF, MF1 and MF2) in response to stress conditions in *Brassica* species.
- Investigating divergence in transcriptional profiles of different copies of genes expressed in response to different stress conditions in *Brassica* species.

The TO1000 annotation was used to assign Bo genes Arabidopsis orthology. The mapping was provided in Supplemental Data Table S7 (Parkin et al., 2014). Each Bo gene mapping to an Arabidopsis gene was assigned to a sub-genome by K_s analysis, in

which the accumulation of mutations (both synonymous, K_s and autonymous, K_a) and deleterious substitutions over time was compared to the ancestral Arabidopsis gene. These sub-genomes have been termed ‘LF’ for the least fractionated with less gene loss, ‘MF1’ for the medium fractionated with moderate gene loss and ‘MF2’ for the most fractionated sub-genome with the most extensive gene loss (Figure 1.4 and Table 1.1a). Genes in *B. oleracea* that were retained in multiple copies in *B. oleracea* (either two or three copies) were subsequently identified as shown in Table 1.1b).

6.2 Results

The *Brassica* genome contains three sub-genomes resulting from two WGDs following the Arabidopsis-Brassica divergence. These sub-genomes show different levels of gene loss following fractionation. The least fractionated (LF) sub-genome corresponds to the least fractionated genome with the highest gene density. The moderately fractionated (MF1) has moderate gene density and the most fractionated sub-genome (MF2) has the lowest gene density (Fig. 1.4 and Table 1.1) (Liu et al., 2014; Parkin et al., 2014). Due to the higher number of genes located on the LF sub-genome, and the fact that fewer of these genes have undergone methylation and accumulation of transposable elements that silence expression (Chen et al., 2015; Cheng et al., 2016; Parkin et al., 2014), it has been proposed that the LF sub-genome shows dominant expression and contributes more greatly to the transcriptome (Cheng et al., 2012a; Parkin et al., 2014). This hypothesis was tested in wild *Brassica* species undergoing salt shock treatment, and in *B. oleracea* GD33DH undergoing salt shock, cold stress and infection with *S. sclerotiorum* by determining the presence of bias in the sub-genome of origin and comparing expression levels (logFC) for differentially expressed genes.

6.2.1 Analysis of sub-genome contribution in *Brassica* species under salt shock

Based on Arabidopsis-*B. oleracea* synteny and the accumulation of synonymous and non-synonymous mutations (used as a proxy for time), the originating sub-genome for each Bo gene with an Arabidopsis ortholog was established (Parkin et al., 2014). These data were used to assess the originating sub-genome of the differentially expressed genes in wild *Brassica* species and GD33DH under various stress conditions. Expression data for a collection of C-genome *Brassica* species responding to salt shock treatment was obtained (Chapter 5). In addition, expression data from GD33DH under salt shock at 24 hpt, obtained as part of a time-series experiment was used (Chapter 4). Detailed of differentially expressed genes in all lines are shown in Table 6.1 and 6.2.

Line	Total DE	LF	MF1	MF2	χ^2	p value
C13001	1,853	616 (600)	452 (421)	318 (361)	7.93	0.018
C13013	5,479	1,712 (1,720)	1,273 (1,208)	959 (1,037)	5.70	0.055
DHLS150	4,278	1,421 (1,420)	1,069 (997)	789 (856)	10.53	0.005
GD33DH	6,558	2,078 (2,150)	1,651 (1,509)	1,237 (1,296)	17.43	1×10^{-04}

Table 6.1: Sub-genome origin of differentially expressed genes in *Brassica* lines following salt shock treatment

The number of genes originating from the LF, MF1 and MF2 sub-genome in lines C13001, C13013, DHLS150 and GD33DH, expected values are given in parenthesis. N/A indicates Bol Genes without an Arabidopsis ortholog. Orthology was assigned using the TO1000 *B. oleracea* genome annotation (Parkin et al., 2014). The critical value (5% significance) for $\chi^2_{d.f=2} = 5.991$.

Of the differentially expressed genes from each line, genes originating from the LF genome were more abundant than the MF1 and MF2 sub-genomes (Table 6.1). Statistical tests were carried out to determine whether this was due to the fact that more genes from the LF sub-genome were retained following biased fractionation of the sub-genomes or due to the preferential expression of genes from LF sub-genome, indicating that this sub-genome plays a dominant role in the stress response. A chi squared test was carried out with the null hypothesis of there is no deviation from the expected proportion of differentially expressed genes originating from the LF:MF1:MF2 (0.433:0.304:0.261) sub-genomes (Parkin et al., 2014). Working with the proportion of genes expressed in each sub-genome effectively normalises for the number of genes present within each sub-genome, as each genome contains a differing number of genes a direct comparison of the number of differentially expressed genes would not be appropriate. For lines examined, there was a significant deviation in the number of differentially expressed genes from the expected proportion in all but the C13013 line ($p < 0.05$, Table 6.1). Interestingly, the deviation from the expected ratio was more prevalent in the MF1 and MF2 groups of genes, which had more (MF1) and less (MF2) of the expected number of genes.

This suggests that although there are more genes retained within the LF sub-genome, the LF sub-genome does not play a more prevalent role in the response to salt shock through the expression of a higher proportion of genes originating from this sub-genome. The MF1 sub-genome, however, expressed more genes than expected suggesting that MF1 sub-genome may play a slightly more prevalent role in the response to salt shock than expected.

In order to determine the effects of gene expression in the different sub-genomes in the response to salt shock in *Brassica* species, an ANOVA analysis was carried out on gene expression data to determine if there are differences between the ‘Treatment’ and ‘Sub-genome’ factors. Two groups of genes were examined for each line:

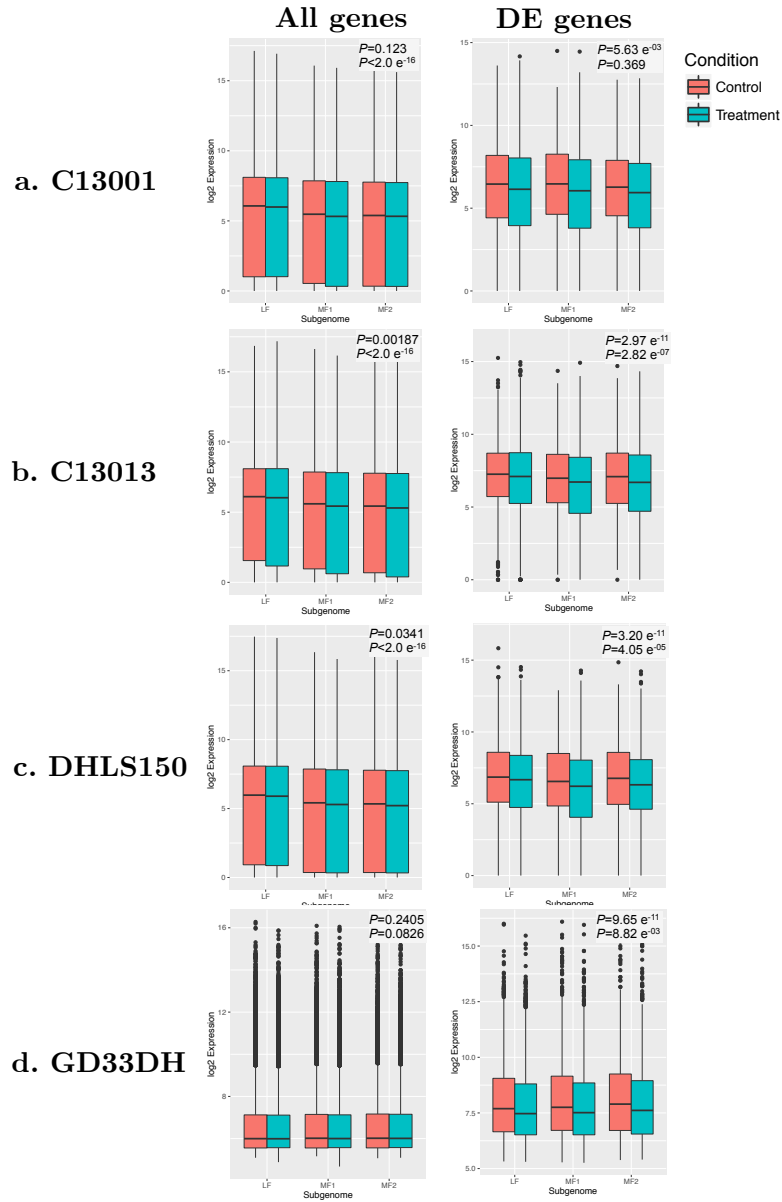


Figure 6.1: Log₂ expression of sub-genome expression in *Brassica* species in response to salt shock

Expression of genes originating from the LF, MF1 and MF2 sub-genome in salt shock conditions in (a) C13001, (b) C13013, (c) DHLS150 and (d) GD33DH. Red is control, blue is treatment, as per the legend. The upper whisker relates to the highest value within 1.5 * of the interquartile range, the lower whisker relates to the lowest value within 1.5 * of the interquartile range. The top line of the box represents the 1st quartile, the middle line is the 2nd quartile (median) and the 3rd quartile of the data is shown by the bottom line of the box. On the *x* axis, expression is log₂ normalised read counts (a - c) or log₂ predicted expression of microarray intensity (d). The sub-genome of origin is on the *y*-axis. The top *p* value relates to the ‘treatment’ term i.e. comparing treatment and control expression. The second *p* value relates to the ‘sub-genome’ term, in which expression between sub-genomes is compared for differences.

Line	Total DE	LF	MF1	MF2	χ^2	p value
Cold	8,152	2,594 (2,587)	1,890 (1,816)	1,490 (1,559)	11.94	0.047
High salt	1,759	618 (609)	449 (428)	340 (367)	2.81	0.201
<i>S. sclerotiorum</i>	160	53 (45)	27 (32)	24 (27)	0.208	0.294

Table 6.2: Sub-genome origin of differentially expressed genes in *B. oleracea* GD33DH following different stress treatment

The number of genes originating from the LF, MF1 and MF2 sub-genome in GD33DH under cold stress, infection with *S. sclerotiorum* and under high salt shock. N/A indicates Bo gene IDs without an Arabidopsis ortholog. The critical value (5% significance) for $\chi^2_{d.f=2} = 5.991$.

1. All genes showing any level of expression in treatment and/or control samples - ‘All genes’. For GD33DH, this included all transcripts with a Bo gene ID, with expression above background level in at least one condition.
2. Expression of differentially expressed genes in treatment and control samples - ‘DE genes’. For the GD33DH line, the \log_2 expression all genes found to be differentially expressed at 24 hpt was used to keep sampling times consistent with the other lines.

Based on the \log_2 expression of genes each group, it was shown that overall there was a significant dominance of gene expression from the LF genome in *Brassica* under salt shock conditions (Fig. 6.1).

The ‘all genes’ group revealed a significant dominance in \log_2 expression of the LF sub-genome in C13001, C13013 and DHLS150 ($p = < e^{-16}$), however, this was not seen in GD33DH (Fig. 6.1 left) possibly because of the differences in technologies used to measure gene expression. This also highlights the higher dynamic range of the microarrays and the ability to detect expression of lowly expressed genes such as TFs.

When the ‘DE genes’ group was considered, a significant difference was seen due to treatment in all cases and in sub-genome expression in C13013 ($p = 2.82 \times 10^{-07}$), DHLS150 ($p = 4.05 \times 10^{-05}$) and GD33DH ($p = 8.82 \times 10^{-03}$; Fig. 6.1 right). That C13001 did not show significant dominance of the LF sub-genome could be because of sample size, as C13001 had substantially fewer differentially expressed genes compared to GD33DH, C13013 and DHLS150 (Table 5.3).

In summary, genes expressed from the LF sub-genome generally show higher expression levels, both in all expressed genes and in differentially expressed genes in *Brassica* species.

6.2.2 Analysis of sub-genome contribution under various stress conditions in *B. oleracea* GD33DH

RNAseq data for GD33DH responding to cold stress (2 °C; sampled 24 hpt; $n=3$), infection with *S. sclerotiorum* (sampled 24 hpt; $n=3$) and high salt shock treatment

(500mM; sampled at 1 hpt; $n=3$) was obtained (Chapter 3) and differentially expressed genes were identified. Under cold stress conditions, GD33DH resulted in 8,152 differentially expressed genes, high salt treatment (500mM) resulted in 1,759 differentially expressed genes and infection with *S. sclerotiorum* resulted in 160 differentially expressed genes (Table 6.2).

In order to determine whether gene expression in GD33DH showed dominant gene expression in the LF sub-genome under different stress conditions, as with other *Brassica* species (Fig. 6.1), the sub-genome origin was determined for genes differentially expressed in cold stress, infection with *S. sclerotiorum* and following high salt shock (Table 6.2).

There was a significant difference from expected in the proportion of differentially expressed genes from the sub-genomes in the cold sample ($p=0.047$), but this was not observed in the high salt treatment or in the infection with *S. sclerotiorum*. This suggests that although there are more genes retained within the LF sub-genome, the LF sub-genome does not play a more predominant role in the response to salt shock through the expression of a higher proportion of genes originating from this sub-genome. An increase in the number of genes expected from the MF1 sub-genome was seen, and fewer genes than expected from MF2, suggesting that MF1 plays a more important role in the response to stress conditions in *Brassica* than would be expected, as seen previously.

In order to determine whether the genes originating from different sub-genomes were expressed at different levels, an ANOVA analysis was carried out on the gene expression data to determine if there are differences between the ‘Treatment’ and ‘Sub-genome’ factors. As above, two groups of genes were examined, ‘All genes’ (Fig. 6.2 left) and ‘DE genes’ (Fig. 6.2 right) for each stress condition.

It was found that in ‘all genes’ group, there was a significant difference in sub-genome expression in each of the stress treatments ($p < 2.0e^{-16}$) with a dominance of expression in the LF sub-genome being evident (Fig. 6.2 left). However, when the ‘DE genes’ group was analysed, this effect was not significant (Fig. 6.2 right) in any of the stress conditions suggesting that the expression dominance of the LF sub-genome is not a feature of the stress response in *Brassica* species.

6.3 Investigating the importance of genes expressed in multiple copies in the transcriptional response of *Brassica* under stress conditions

The *Brassica* genome contains three sub-genomes which show different levels of gene loss following fractionation (Fig. 1.4). Some genes such as clock genes and TFs are preferentially retained across all three sub-genomes (Blanc and Wolfe, 2004; Lou et al.,

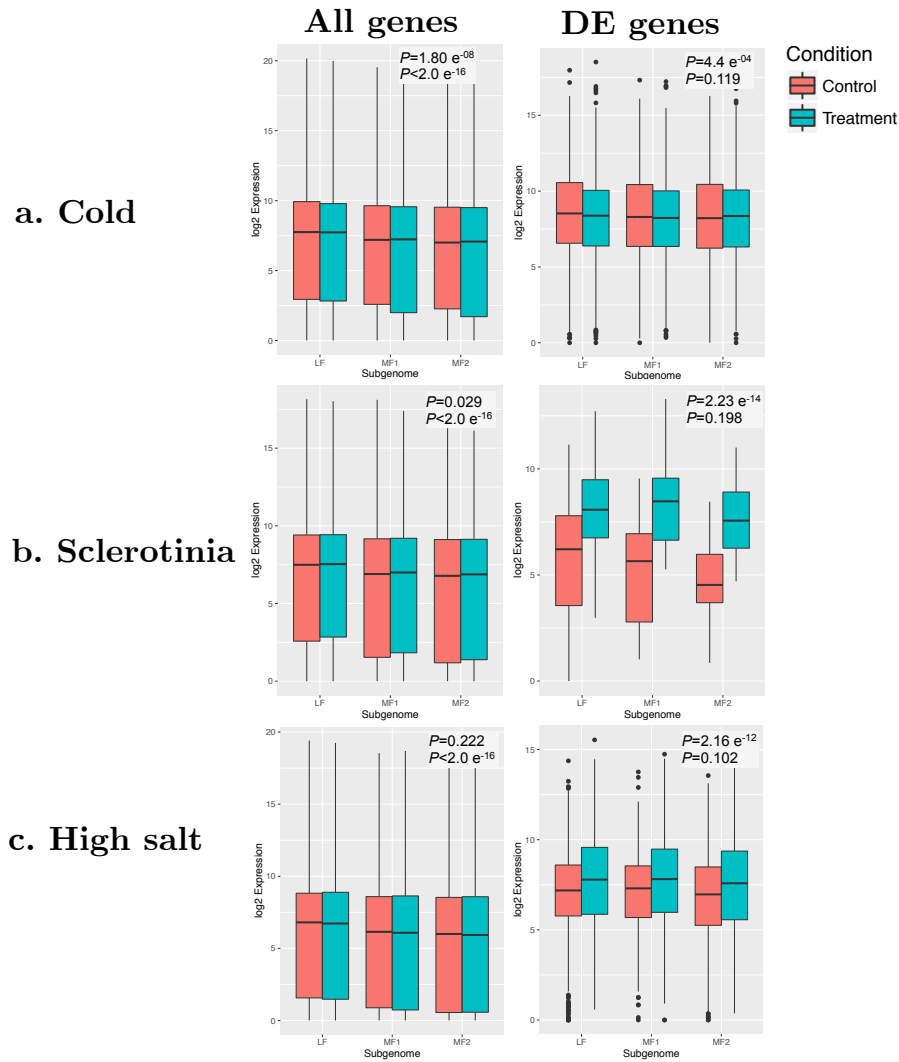


Figure 6.2: Log₂ expression of sub-genome expression in *B. oleracea* GD33DH following different stress treatment

Expression of genes originating from the LF, MF1 and MF2 sub-genome in GD33DH under (a) cold stress, (b) infection with *S. sclerotiorum* and (c) under high salt shock. Red is control, blue is treatment, as per the legend. The upper whisker relates to the highest value within 1.5 * of the interquartile range, the lower whisker relates to the lowest value within 1.5 * of the interquartile range. The top line of the box represents the 1st quartile, the middle line is the 2nd quartile (median) and the 3rd quartile of the data is shown by the bottom line of the box. On the *x* axis, expression is log₂ normalised read counts. The sub-genome of origin is on the *y*-axis. The top *p* value relates to the ‘treatment’ term i.e. comparing treatment and control expression. The second *p* value relates to the ‘sub-genome’ term, in which expression between sub-genomes is compared for differences.

Line	1 copy	2 copies	3 copies	No ortholog
C13001	1,108	121	10	467
C13013	2,869	453	67	1,535
DHLS150	2,380	366	55	1,027
GD33DH	3,176	478	49	2,279

Table 6.3: The number of multiple-copy genes in the differentially expressed genes of the *Brassica* lines under salt shock

The number of paralogous gene groups (1-3 copies) that are orthologous to a single copy of an Arabidopsis gene, in the differentially expressed genes of lines C13001, C13013, DHLS150 and GD33DH following salt shock treatment. ‘No ortholog’ indicates Bol genes without an Arabidopsis ortholog. Orthology was assigned using the TO1000 *B. oleracea* genome annotation (Parkin et al., 2014).

2012), whilst other groups of genes are preferentially retained as a single copy to not disturb the stoichiometric balance of the cellular machinery. These genes are generally involved in essential housekeeping or form parts of large dose-sensitive protein complexes (De Smet et al., 2013; Schnable et al., 2012).

The expression of genes in multiple copies was investigated to determine whether increased copy number had an effect on the transcriptome in *Brassica* species under various stress conditions.

6.3.1 Paralogous genes expressed in response to salt shock in *Brassica* species

In the following section, expression of the number of paralogs that were differentially expressed in response to stress, regardless of the number retained in the *B. oleracea* TO1000 genome, was determined as shown in Table 6.3. For instance, genes that expressed in a single copy may in fact be retained in two or three copies in the genome, however only one shows differential expression in this experiment.

Across all of the lines, the majority of genes were expressed in a single copy and the number of differentially expressed paralogs decreased as copy number increases suggesting differential expression of all three paralogs (‘triplets’) is a rare event. In the C13001 line, there are only 10 differentially expressed triplets, increasing to 67 triplets for the C13013 line (Table 6.3).

Of the genes differentially expressed in triplicate, a greater proportion (10.99%) were TFs compared to those differentially expressed in duplet (8.76%) or as single copies (8.64%). TFs differentially expressed in triplicate across the wild C-genome *Brassica* and GD33DH included *MYB96*, telomeric DNA binding protein 1 (*TBP1*), GATA transcription factor 17 (*GATA17*), response regulator 4 (*ARR4*), WRKY DNA-binding protein 18 (*WRKY18*), *JAZ1* amongst others.

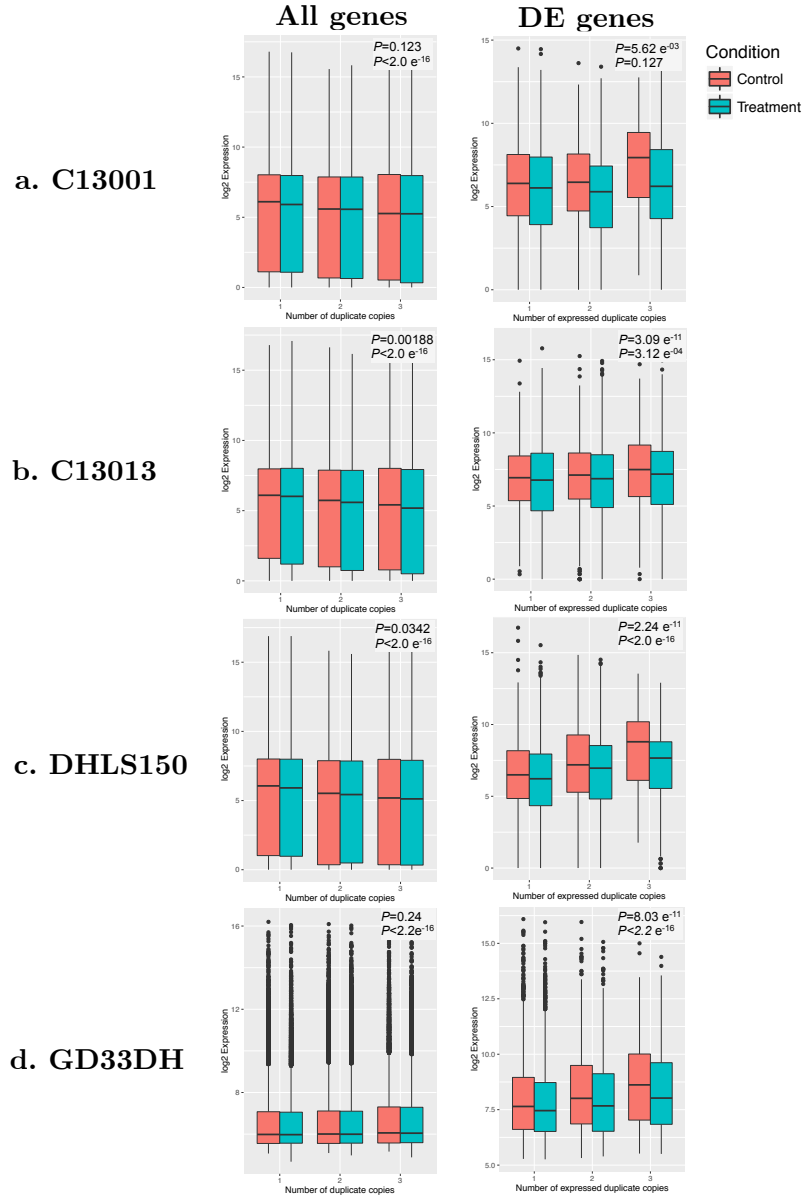


Figure 6.3: Log₂ expression of genes expressed in multiple copies in *Brassica* species in response to salt shock

Red is control, blue is treatment, as per the legend. The upper whisker relates to the highest value within 1.5 * of the interquartile range, the lower whisker relates to the lowest value within 1.5 * of the interquartile range. The top line of the box represents the 1st quartile, the middle line is the 2nd quartile (median) and the 3rd quartile of the data is shown by the bottom line of the box. On the *x* axis, expression is log₂ normalised read counts (C13001, C13013 and DHLS150) or log₂ predicted expression of microarray intensity. Copy number of paralogs is on the *y*-axis. The top *p* value relates to the ‘treatment’ term i.e. comparing treatment and control expression. The second *p* value relates to the ‘paralog’ term, in which expression between genes expressed in different copy numbers is compared for differences.

Stress	1 copy	2 copies	3 copies	No ortholog
Cold	4,091	772	113	2,178
<i>S. sclerotiorum</i>	92	6	0	56
High salt	1,139	113	14	352

Table 6.4: The number of multiple-copy genes in the differentially expressed in GD33DH under different stress conditions

The number of paralogous gene groups (1-3 copies) that are orthologous to a single copy of an Arabidopsis gene in the differentially expressed genes in GD33DH under cold stress, infection with *S. sclerotiorum* and high salt shock. ‘No ortholog’ indicates Bo gene IDs without an Arabidopsis ortholog. Orthology was assigned using the TO1000 *B. oleracea* genome annotation (Parkin et al., 2014).

In order to determine the whether paralogs expressed in multiple copies are expressed at a different level compared to genes that are expressed as single copies, an ANOVA analysis was carried out on gene expression data to determine if there are differences between the ‘Treatment’ and ‘Duplicates expressed’ factors. As above, two groups of genes were examined, ‘All genes’ and ‘DE genes’ for each line (Fig. 6.3).

In the ‘all genes’ group it was shown that there was a significant difference in spread of the \log_2 expression of genes expressed in multiple copies in all lines; C13001, C13013, DHLS150 and GD33DH ($p < 2.0e^{-16}$; Fig. 6.3, left). In this group of genes it was seen that the \log_2 expression of gene expressed in a single copy is higher than that of genes expressed in two or more copies. This was contrary to the expression pattern seen in the ‘DE gene’ group of genes, where the \log_2 expression increased as the number of copies expressed increased. This effect was significant in C13013 ($p = 3.12e^{-04}$), DHLS150 ($p < 2.0e^{-16}$) and GD33DH ($p = 2.0e^{-16}$) lines (Fig. 6.3, right), but not significant in C13001. The number of differentially expressed genes expressed in multiple copies in this line was considerably lower (Table 6.3). These results suggest that as the number of paralogs expressed increases, so does the level of expression. This effect has been seen previously in *B. rapa* (Schnable et al., 2012), *Caenorhabditis elegans* and human (Padawer et al., 2012) and in *Phaseolus vulgaris* (mesoamerican common bean, Vlasova et al., 2016), where the average expression of genes in multiple copies tended to be significantly higher than genes expressed in single copies.

6.3.2 Paralogous gene expression in other stress conditions in *B. oleracea* GD33DH

The paralogous groups of genes differentially expressed in response to different stress conditions in GD33DH, identified using Ara-Bol orthology (Parkin et al., 2014) in the TO1000 genome, are shown in Table 6.4. Across all stress conditions, the majority of

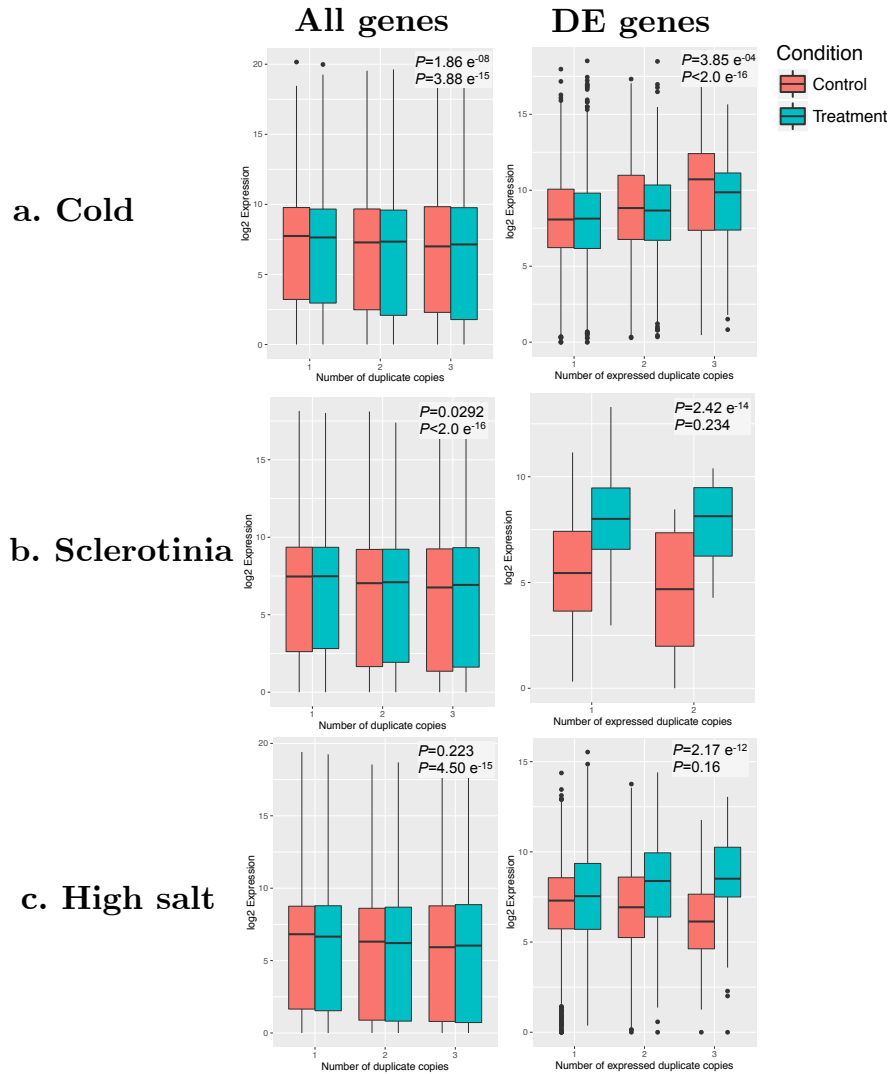


Figure 6.4: Log₂ expression of genes expressed in multiple copies in *B. oleracea* GD33DH in response to stress

Expression of paralogous gene sets (1-3 copies) that are orthologous to a single copy of an Arabidopsis gene in the differentially expressed genes in GD33DH under cold stress, infection with *S.sclerotiorum* and high salt shock. Red is control, blue is treatment, as per the legend. The upper whisker relates to the highest value within 1.5 * of the interquartile range, the lower whisker relates to the lowest value within 1.5 * of the interquartile range. The top line of the box represents the 1st quartile, the middle line is the 2nd quartile (median) and the 3rd quartile of the data is shown by the bottom line of the box. On the *x* axis, expression is log₂ normalised read counts. Copy number of paralogs is on the *y*-axis. The top *p* value relates to the ‘treatment’ term i.e. comparing treatment and control expression. The second *p* value relates to the ‘paralog’ term, in which expression between genes expressed in different copy numbers is compared for differences.

genes were expressed as a single copy and the number of differentially expressed paralogs decreased as copy number increases suggesting differential expression of all three paralogs ('triplets') is a rare event. Cold stress exhibited the largest number of triplets (113 groups of triplets) in the experiment, whilst infection with *S. sclerotiorum* did not have any differentially expressed genes in triplicate, though this is due to the small sample size (Table 6.4).

In order to determine the whether paralogs expressed in multiple copies are expressed at a different level compared to genes that are expressed as single copies, an ANOVA analysis was carried out on gene expression data to determine if there are differences between the 'Treatment' and 'Duplicates expressed' factors. As above, two groups of genes were examined, 'All genes' and 'DE genes' for each line (Fig. 6.3).

In the 'all genes' group it was shown that there was a significant difference in spread of the \log_2 expression of genes expressed in multiple copies in all lines; cold stress ($p = 3.88 \times 10^{-15}$), infection with *S. sclerotiorum* ($p < 2.0 \times 10^{-16}$) and high salt shock ($p = 4.50 \times 10^{-15}$) (Fig. 6.4, left).

When the 'DE gene' group was examined, this effect is lost in the infection with *S. sclerotiorum* and high salt shock ($P > 0.05$), most likely due to the small sample size of differentially expressed genes being unable to provide a robust estimation of variation for these lines (Table 6.4). However, in cold stress there is a significant difference in paralog expression ($p < 2.0 \times 10^{-16}$), the sample size in this group is sufficient to tease apart variation cause by paralog number (Fig. 6.4, right). Genes that are expressed in multiple copies show a significantly higher expression, as seen above.

6.4 The transcriptional fate of genes expressed in multiple copies

There are several models that have been proposed to explain the fate of genes arising from WGD events (as discussed in Chapter 1). The sub-/neo-functionalization model suggests that function and expression is either partitioned between paralogs ('sub-functionalization'), functional diversification of one paralog ('neo-functionalization') or one paralog may experience loss of function ('pseudogenization') (Blanc and Wolfe, 2004; Cusack and Wolfe, 2006; Force et al., 1999; Freeling et al., 2015; Ohno, 1970; Schnable et al., 2012).

To investigate the fate of genes expressed in response to salt shock in *Brassica* species, the patterns of expression exhibited by genes with two and three expressed copies were analysed by comparing logFC between paralogs (Fig. 6.5). The ratio of logFC between the maximally expressed paralog and minimally expressed paralog was calculated.

To categorize the fate of the differentially expressed paralogs:

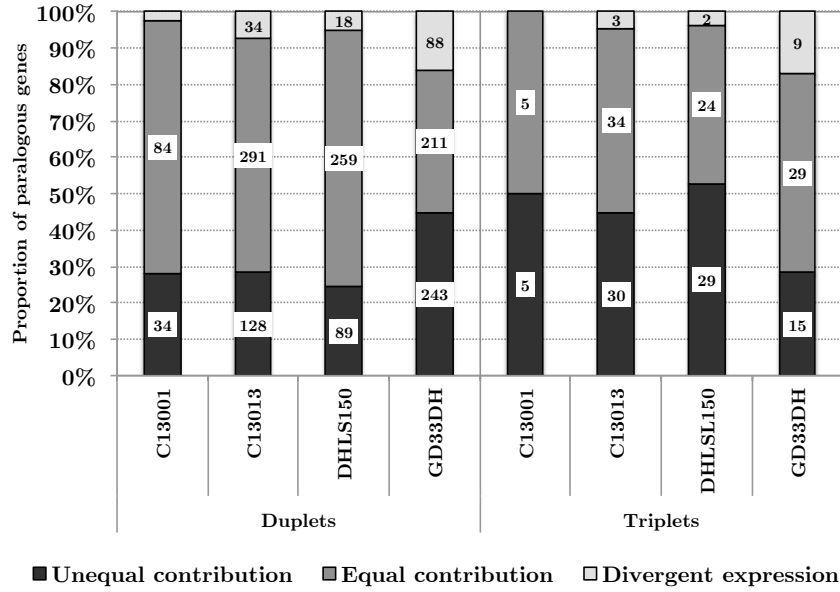


Figure 6.5: Gene fate of 2-3 copy paralogous genes in *Brassica* species

The proportion of 2-3 copy differentially expressed paralogous genes experience different fates. Paralogs contributing to expression in an equal manner ($\text{LogFC}_{\max}/\text{LogFC}_{\min} \leq 1.5$; indicated in dark grey), an unequal manner ($\text{LogFC}_{\max}/\text{LogFC}_{\min} \geq 1.5$; indicated in mid-grey) or with divergent expression (LogFC not occurring in the same direction; indicated in light grey). The number of genes present in each group is given in the chart.

- Sub-functionalization - could be either equal or unequal. Paralogs were considered to be expressed in an ‘equal’ manner if within each paralogous group $\text{LogFC}_{\max}/\text{LogFC}_{\min} \leq 1.5$. An ‘unequal’ expression of paralogs in the same direction was considered to have occurred if $\text{LogFC}_{\max}/\text{LogFC}_{\min} \geq 1.5$, indicating dominance or partial redundancy of an individual paralog (Qian et al., 2010).
- Neofunctionalization - ‘divergent’ expression in which paralogs were expressed in different directions.

6.4.1 The fate of genes expressed in multiple copies in *Brassica* species in response to salt shock

The results of this analysis are shown in Figure 6.5, where it can be seen that sub-functionalization with equal contribution of expression is the primary fate of most paralogs expressed in duplicate (70%) and sub-functionalization with a similar proportion of equal or unequal expression patterns shown in triplicate. It was found that most genes were differentially expressed in the same direction, however, a small number of paralogs were expressed in different directions within their paralogous groups indicating transcriptional divergence of these paralogs (Fig. 6.5). It was previously shown that as the number of

paralogs increases, so does the average expression. Here it can be shown that as the number of paralogs increases, the more unequal the expression of paralogs becomes.

Using the GD33DH time-course data (described in Chapter 4), it is possible to examine expression profiles of some of the genes expressed as triplets, illustrating the different fates of paralogs described above (Fig. 6.5).

Triplets A and B encode 60S (L8) and 30S (S9) ribosomal proteins, respectively (Fig. 6.6a and b) and are examples of triplets with a fate of equal sub-functionalization, in which all genes are down-regulated at the same rate in response to salt shock. This is the most common fate of paralog expression in lines C13001, C13013 and DHLS150.

Triplets C (Cytochrome C oxidase 6B) and D (ABA receptor *PYR1*) (Fig. 6.6c and d) represent unequal partitioning of expression between down-regulated paralogs (sub-functionalization) and triplets E and F (Fig. 6.6e and f) are examples of unequal sub-functionalization of up-regulated paralogs. This unequal expression of paralogs is more common as the number of paralogs expressed increases (Fig. 6.5), possibly due to the relationship between copy number and expression as shown in Figure 6.3. According to the gene dosage hypothesis, paralogs that encode proteins involved in large multi-protein subunits such as ribosomal subunits and complex regulatory networks such as the ABA receptor *PYR1* are more likely to be retained in multiple copies (Birchler and Veitia, 2007). That expression between these paralogs does not diverge suggests that all are important within the transcriptional network that they belong to, and therefore are co-regulated.

Finally, triplets G and H (Fig. 6.6g and h) are possible examples of neo-functionalization, paralogs which are differentially expressed in opposing directions during the stress response. As shown in this analysis, neo-functionalization of paralogs is the less common fate of paralogs (Fig. 6.5).

6.4.2 The fate of genes retained in multiple copies in GD33DH in response to stress conditions

To investigate the fate of genes expressed in response to salt shock in *Brassica* species, the patterns of expression exhibit by genes with two and three expressed copies were analysed by comparing logFC between paralogs (Fig. 6.5). The ratio of logFC between the maximally expressed paralog and minimally expressed paralog was calculated and using the criteria described above, each duplet or triplet was assigned a category - sub-functionalization (in which $\text{LogFC}_{\text{max}}/\text{LogFC}_{\text{min}} \leq 1.5$ indicates equal expression or $\text{LogFC}_{\text{max}}/\text{LogFC}_{\text{min}} \geq 1.5$ indicates unequal expression) or neofunctionalization (LogFC occurring in opposing directions within the paralogous group).

The results of this analysis are shown in Figure 6.7, where it can be seen that sub-functionalization with equal contribution of expression is the primary fate of most

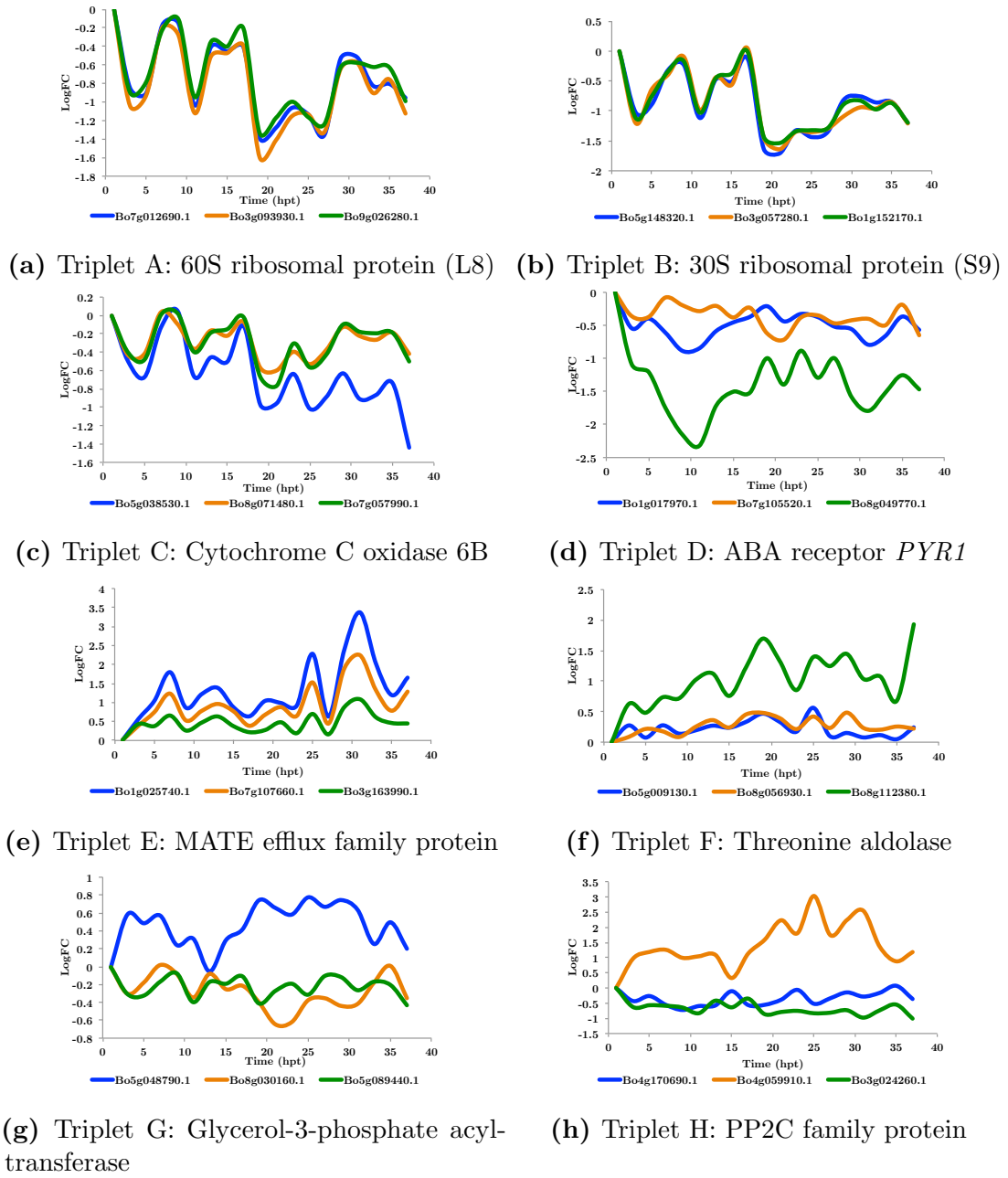


Figure 6.6: Expression profiles of triplet paralogs with different fates during salt shock in GD33DH

LogFC is found on the x -axis, time (hpt) is located on the y -axis. The blue line represents genes originating from the LF sub-genome, MF1 is represented by the orange line and MF2 is the green line.

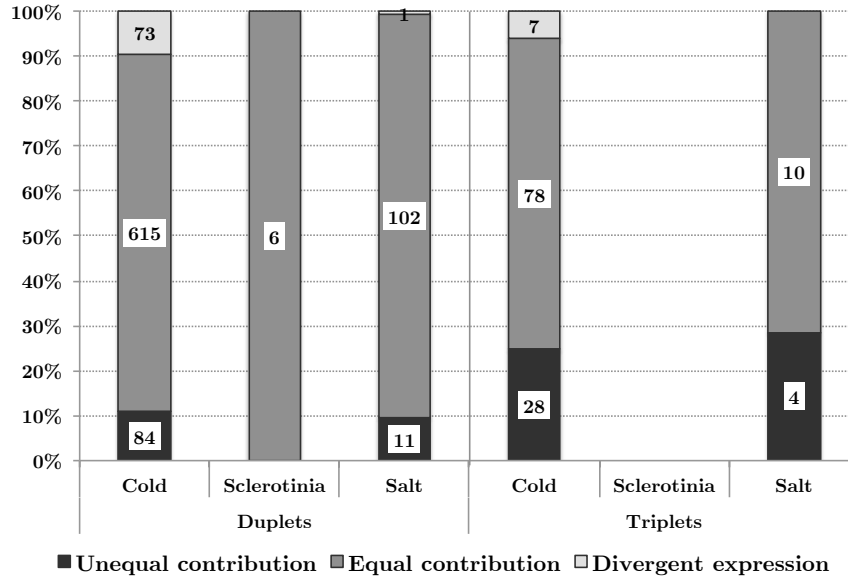


Figure 6.7: Gene fate of 2-3 copy paralogous genes in *Brassica* species

The proportion of 2-3 copy differentially expressed paralogous genes experiencing different fates. Paralogs contributing to expression in an equal manner ($\text{LogFC}_{\text{max}}/\text{LogFC}_{\text{min}} \leq 1.5$; indicated in dark grey), an unequal manner ($\text{LogFC}_{\text{max}}/\text{LogFC}_{\text{min}} \geq 1.5$; indicated in mid-grey) or with divergent expression (LogFC not occurring in the same direction; indicated in light grey). The number of genes present in each group is given in the chart.

paralogs expressed in duplicate ($\sim 70\%$), with a small proportion of genes showing unequal expression of duplicate paralogs (0 - 10%). Due to the small number of differentially expressed genes following infection with *S. sclerotiorum*, there were only 6 duplets, each showing expression in an equal manner. It was found that most genes were differentially expressed in the same direction, however, a small number of paralogs were expressed in different directions within their paralogous groups indicating transcriptional divergence and potential neofunctionalization (Fig. 6.5). There is a similar pattern seen in the triplicate group, however as seen previously (Fig. 6.5), there is a larger proportion of subfunctionalized expression of paralogs with unequal contribution to the total expression.

6.5 Discussion

To investigate the effects of WGD in the recent evolutionary history of *Brassica* species on the stress response, whole transcriptome sequencing was carried out in C-genome *Brassica* species under various stress conditions including salt shock, cold stress and infection with *S. sclerotiorum*. Gene expression was examined in terms of sub-genome origin and expression of paralogs with the same ancestral Arabidopsis gene expressed in multiple copies. Although many studies have focussed on the effects of WGD on retention of genes (Adam et al., 2003; Akama et al., 2014; Cheng et al., 2016; Li et al., 2014a; Lou et al.,

2012; Moghe et al., 2014; Murat et al., 2014), relatively few studies analyse the effects of WGD events on gene expression under stress conditions (Li et al., 2014c; Zhang et al., 2016).

6.5.1 Sub-genome dominance

A dominance effect of the LF sub-genome compared to the MF1 and MF2 sub-genomes was seen in the expression of all genes from different lines of *Brassica* species and in different stress conditions. This effect has been reported previously in *B. rapa* (Cheng et al., 2012a, 2016; Wang et al., 2011b) and *B. oleracea* (Liu et al., 2014; Parkin et al., 2014) and other non *Brassica* species (Hovav et al., 2008; Leach et al., 2014; Murat et al., 2014; Pont et al., 2013). When the DE genes were considered, the evidence of this dominance effect was largely determined by the number of differentially expressed genes, with lines with a large number of salt-induced differentially expressed genes in C13013, DHSL150 and in cold stress in GD33DH showing significant differences in expression by sub-genome. Smaller groups of genes identified in salt shocked C13001 and GD33DH high salt treatment and infection with *S. sclerotiorum* showing no significant difference.

This suggests that genes found on this sub-genome have a marginally higher level of expression compared to genes found on older sub-genomes and could be preferentially selected as potential breeding targets in development of stress tolerant varieties. The effect however is subtle, and when small groups of differentially expressed genes are identified, the difference is undetectable suggesting that the stress-specific transcriptomic response of *Brassica* species to stress conditions is not overwhelmingly caused by preferential use of the LF sub-genome.

In the recent allotetraploid *Coffea arabica* under temperature stress conditions, the differential contribution of the sub-genomes to the response compared to diploid parents was marginal suggesting that the enhanced tolerance seen by the polyploid was not due to the use of homoeologs from a dominant sub-genome (Combes et al., 2013).

6.5.2 Stress-specific expression of genes in multiple copies

The relationship between the number of expressed paralogs and average expression was interesting. When the group of all genes was examined, average expression was higher in genes retained in single copies, rather than multiple copies. Many of these genes were housekeeping genes and genes involved in key cellular functions that were unaffected by stress. This observation was reversed when differentially expressed genes were considered, in which the average expression of genes in multiple copies tended to be significantly higher than genes expressed as single copies.

Increased expression in all genes expressed in multiple copies has been seen in *B. rapa*

(Schnable et al., 2012), *Caenorhabditis elegans* and human (Padawer et al., 2012) and in *Phaseolus vulgaris* (mesoamerican common bean) (Vlasova et al., 2016), though these analyses were not carried out under stress conditions.

This suggests an important role for WGD in the evolution of the stress response in *Brassica*. It supports the hypothesis that on the whole, genes are preferentially retained and expressed as single copies, but stress-responsive genes are more likely to be retained and expressed in multiple copies under stress conditions suggesting that WGD confers an enhanced ability of *Brassica* to adapt to stress conditions.

There is a preference in gene class as to which genes are retained in multiple copies in auto- and allo-polyploid species, such as circadian clock genes (Lou et al., 2012). TFs (Blanc and Wolfe, 2004) and ribosomal proteins (Wang et al., 2011b) and genes involved in environmental adaptivity (Blanc and Wolfe, 2004; Ha et al., 2009). The proteins encoded for by these genes are generally found in large, highly interactive networks which exhibit stoichiometric balance. The loss of a gene copy would result in an unbalanced protein network, affecting cellular function and downstream gene expression. This is phenomenon is known as the Gene Balance Hypothesis (Birchler and Veitia, 2007).

The fate of retained paralogs varied between pseudogenization, sub-functionalization and neo-functionalization. In the duplets of the differentially expressed genes, the most likely fate of paralogous gene expression was sub-functionalization of expression between paralogs. When the triplets were examined, this proportion shifted to an equal expectation of paralogs having an equal or unequal contribution to expression suggesting that sub-functionalization is the main fate of genes that are salt-responsive. In all groups, potential neo-functionalization, defined by genes with expression in different directions was not the primary fate of genes possibly due to the fact that genes present in multiple copies as a result of a WGD event tend to diverge more slowly than other modes of duplication (Wang et al., 2011c). In keeping with the gene balance hypothesis, neo-functionalization of a gene within a large protein subunit would have a detrimental effect on the stoichiometric balance and would be actively selected against, possibly accounting for the low incidence of neo-functionalization in the differentially expressed genes (Birchler and Veitia, 2012).

Additionally, there was a large number of genes differentially expressed in fewer copies than were retained on the genome (e.g. duplets) suggesting potential pseudogenization, or unequal contribution of one paralog to expression. This process can be seen in action when examining the expression profiles of genes in which one copy has reduced expression compared to the others (Fig. 6.6). Similar observations of gene fate following WGD have been made in *Paramecium* sp and yeast (Gout and Lynch, 2015), in which expression of one gene copy is decreased compared to others eventually leading to pseudogenization of this gene copy. Several gene families have been characterised in the *Brassica* genus and have been found to show evidence of pseudogenization of multi-copy genes including the

ARF gene family (Mun et al., 2012) and the SnRK2 gene family (Huang et al., 2015) in both in *B. rapa*.

6.6 Chapter summary

Brassica species have undergone three rounds of WGD in the recent evolutionary history, resulting in a highly complex genome (Cheng et al., 2014; Parkin et al., 2014; Wang et al., 2011b). Whilst there is a dominance of the LF sub-genome in global gene expression, the effect on the stress-responsive gene expression potentially favours the MF1 sub-genome, contrary to the literature. Genes expressed in multiple copies show a lower level of global gene expression. However, under stress conditions, genes differentially expressed in multiple copies were expressed at higher levels compared to single copy differentially expressed genes, suggesting that WGD and the availability of multiple copies of stress responsive genes is an important mechanism in the evolutionary process, allowing for increased adaptability to environmental stress conditions.

Chapter 7

General discussion

This thesis reports the development and use of transcriptomic technologies to elucidate biological mechanisms associated with the response to salt shock in C-genome *Brassica* species. A high-resolution time-series profile of every transcript present on a newly designed C-genome *Brassica* array has been produced using techniques minimising technical and experimental variation that is often associated with two colour microarrays. Transcripts with altered expression over the time-series have been identified and grouped into clusters based on the time at which they were first differentially expressed in order to predict biological function and also by the shape of expression profile to predict potential transcriptional regulators.

Significant variation in tolerance to salt shock was found in wild C-genome *Brassica* species, demonstrating the potential of crop wild relatives as a source of germplasm in the development of stress tolerant crop plants. RNAseq analysis of tolerant and susceptible lines allowed elucidation of transcriptomic changes that may cause a different physiology leading to enhanced tolerance to salt shock. Finally, by considering the evolutionary history of *Brassica* species, it has been possible to determine aspects of genome architecture are likely to be important in the response to salt shock. The recent availability of the *B. oleracea* TO1000 genome has allowed the global, in depth study previously reserved for model organisms such as *Arabidopsis* to be carried out in a non-model crop species.

This thesis highlights the importance of translating knowledge, both biological and methodological, from model organisms into crop plants. This will have an impact on the development of stress resilient crop plants that will be capable of sustaining a growing population amid the effects of climate change.

7.1 Development of transcriptomic resources in *Brassica oleracea*

Chapter 3 describes the development of a new Agilent microarray aimed at C-genome *Brassica* species using recent sequence data from the *B. oleracea* genome (Parkin et al., 2014), leaf and root RNAseq data (generated as part of the VeGIN project) and stress RNAseq data in GD33DH and C-genome *Brassica* (this thesis).

The microarray is comprised of 160,324 60-mer probes allowing for the distinction of 53,387 *B. oleracea* TO1000 genes (representing 90.1% of currently available gene models). By combining genomic sequence with RNAseq sequence information an extra dimension to the array is added, allowing study of 32,770 unannotated transcripts and also, where many probes map to an individual gene, allowing expression measurements for different sections of that gene. This means the array will be useful for the analysis of novel transcripts and alternative splicing events in C-genome *Brassica* genes.

A GO annotation of the TO1000 genome functional analysis of gene expression was previously unavailable. GO annotation for 43,190 *B. oleracea* TO1000 genes was generated using orthology to a selection of Brassicaceae sequences downloaded from the NCBI's *nr* database and through comparative analysis of protein domains present in each gene.

The newly generated microarray, together with the GO annotation for the *B. oleracea* TO1000 genome adds to the excellent genomic resources available for C-genome *Brassica* species (Liu et al., 2014; Love et al., 2010; Parkin et al., 2014; Trick et al., 2009). Use of the array and associated annotation information will allow further transcriptomic study and subsequent downstream analysis for this important crop species.

7.2 High-resolution time series transcriptomics of salt shock in *Brassica oleracea*

Using the *Brassica* C-genome microarray developed in Chapter 3, a novel dataset was collected in Chapter 4 which comprised of a high-resolution time series in which global transcriptomic measurements were made every 2 h for 36 h following salt shock, capturing subtle fluctuations in transcript levels for the 160,324 transcripts present on the array. The experiment was designed to reduce technical and biological variation between measurements and to make robust comparisons over the time points. This was achieved by using a complex loop design and the use of MAANOVA which was adapted locally (McHattie, 2011) to extract predicted expression values for each time point, and for the estimation of differentially expressed transcripts. Two methods of testing for differential expression were used – *F*-tests in MAANOVA and a Gaussian Process Two Sample test providing

confidence in the resulting 11,754 differentially expressed transcripts. A high proportion of the *B. oleracea* transcriptome was altered under salt shock and observations on the response of *B. oleracea* GD33DH to salt shock could be made. From this, it was seen that changes in gene expression occurred in two waves following salt shock and a chronology of the first 36 hpt of the stress response was established. This revealed distinct changes in transcript function at specific time points. For example, a vast reduction in gene expression following salt shock indicated that metabolism is stalled while the plant makes key adjustments to rebalance cellular homeostasis. Ultimately, growth is affected which has a negative effect on crop yield.

The novelty of this experiment is that a dataset of this high-resolution has never been collected in a non-model organism. Several similar datasets were collected in *Arabidopsis* under various stress conditions (Bechtold et al., 2016; Breeze et al., 2011; Lewis et al., 2015; Windram et al., 2012). This highlights the usefulness of transcriptomic technologies and the development of data analysis methods in *Arabidopsis* for the translation from model to crop.

As touched upon in this thesis, but not explored to its full potential, the data produced by such a dataset is highly amenable to modelling gene regulatory networks under stress conditions using network inference algorithms such as CSI (Penfold et al., 2015; Penfold and Wild, 2011). This would result in the identification of genes whose expression profiles which may have the largest influence on down-stream profiles. Due to the strong diurnal nature of many expression profiles masking causal structure, it would be necessary to negate these strong time of day effects prior to modelling.

Another interesting aspect to be developed following on from the results of this chapter would be the use of mutants, particularly altered expression mutants for genes that were identified as potential regulators of the salt shock response, such as *KNAT3*, *CRF2*, *HB6*, *WRKY28* etc. This could be carried out either in *Arabidopsis* using SALK T-DNA insertion lines or by developing mutant *B. oleracea* GD33DH lines using CRISPR technology (Bortesi and Fischer, 2015) to disrupt the gene(s) of interest.

Given that climate change will result in unpredictability of many different types of stress conditions, such as greater prevalence of pests, altered atmospheric CO₂ and increased temperature, it is important to understand the key genes involved in the response of plants to multiple stresses in order to develop lines with a broad stress tolerance. For this, it would be crucial to assess transcriptomic changes under different stress conditions, as well as over a longer time period over days rather than hours.

7.3 Analysis of gene expression in response to salt shock in wild C-genome *Brassica* species

The presence of genetic diversity in germplasm is highly important in the development of stress tolerant crop species if yields are to be maintained following the environmental impact of climate change. The required variation is not always present in commercially developed lines, since many generations of selective breeding will probably have narrowed the gene pool.

Following the identification of differentially expressed transcripts during the first 36 h of the salt shock response in GD33DH, developed from the commercial line ‘Green Duke’ (Chapter 4), it was decided to investigate variation in salt tolerance as well as associated gene expression in a collection of wild C-genome *Brassica* species (Chapter 5).

As demonstrated in Chapter 5, there is clear genetic variation in response to salt shock shown in the C-genome DFFS collection, making the population ideal germplasm for the development of stress tolerant C-genome *Brassica*. This population has been screened previously for diversity in shoot Ca^{2+} and Mg^{2+} concentration (Broadley et al., 2008; White et al., 2010), seed oil content (Barker et al., 2007) and water use efficiency (Thompson, 2011, 2009). Following the salt shock phenotype screen which measured physiological traits such as height, fresh/dry weight and leaf area several lines with tolerance to high salinity were identified, including C10128 (*B. oleracea*) and C13013 (*B. incana*).

RNAseq analysis of one of the potential highly tolerant lines (C10128), which showed a highly efficient, minimal transcriptomic response to salt shock. This line was shown to maintain a high $\text{K}^+:\text{Na}^+$ ratio under salt shock conditions. Such was the tolerance of this line, that there were very few differentially expressed genes at 24 hpt, and there was virtually no effect seen on plant height 14 days post treatment. In the other two sequenced lines, it was again observed that there were more down-regulated genes compared to up-regulated genes. Some of these genes overlapped, possibly identifying a ‘core’ response to salt shock. By comparing contrasting genotypes, it could be predicted that the enhanced tolerance may have originated from altered regulation of genes involved in metabolism and growth in the tolerant line early in the salt shock response. By down-regulating many aspects of metabolism and growth such as primary metabolism, photosynthesis and cell growth, the plant is able to redirect energies into negating further damage cause by high salt conditions.

Further RNAseq analysis to clarify mechanisms of tolerance to salt shock, would be beneficial. In Chapter 4, two phases of the response to salt shock were identified in GD33DH at 0 - 16 hpt (phase 1) and 18 - 36 hpt (phase 2) which showed differences in the number of differentially expressed transcripts and also in biological function. It would be interesting to sequence additional time points in the wild *Brassica* lines for comparison

with GD33DH chronology. Throughout the experiments presented in Chapters 4 and 5, leaf material from fully expanded leaves were used for expression profiling. It would be interesting to see the effects of salt shock on younger leaves, which were not fully expanded at the time of salt shock and also the transcriptomic changes seen in the roots. Finally, to further characterise and identify potential mechanisms enabling salt tolerance it would be necessary to sequence a larger number of lines of differing tolerance levels and levels of domestication.

A Genome Wide Association Study (GWAS) using a panel of SNPs that were developed as part of the Defra funded VeGIN project might be useful to assess allelic diversity in the wild C-genome *Brassica* that were studied in this thesis. Association of a salt tolerant phenotype with the presence/absence of individual SNPs might be used to identify molecular markers that would be useful for breeders in developing new varieties.

7.4 Transcriptional divergence of stress responsive paralogs in C-genome *Brassica* species

The impact of WGD events in the evolutionary history of *Brassica* species was examined in Chapter 6 where it was found that there was no strong effect of sub-genome dominance in stress responsive genes in C-genome *Brassica* species. However, higher expression of genes differentially expressed in multiple copies compared to single copies was observed. This was contrary to global gene expression, which saw higher levels of expression of genes expressed as a single copy. Stress responsive genes such as TFs etc. are more likely to be involved in large multi-protein subunits, where as housekeeping genes such as DNA repair genes tend to work as single entities, these results support the gene dosage hypothesis. WGD events resulting in differential expression of multiple paralogs of stress responsive genes such as TFs allow for the fine tuning of the stress response and may be important in the evolution of stress tolerance.

Further work following from the results obtained in this Chapter could involve the use of new genome editing technologies such as CRISPR to disrupt one (or more) of the gene paralogs and determining the effects on the resulting gene regulatory networks.

Appendix A

MAANOVA analysis script

This R script is the top-level script for the microarray analysis of the senescence data. It calls a number of functions which operate MAANOVA's pre-ANOVA quality control, ANOVA model fitting and post-ANOVA identification of significantly differentially expressed genes.

```
library(whrimaanova)

# Change to directory containing all the data
setwd('/home/christinehicks/maanova')

# Read in the data
man_data <- read.madata('noBGcorrection_datafile.txt',
                        'design.txt',
                        arrayType='twoColor',
                        log.trans=TRUE,
                        spotflag=FALSE,
                        probeid = 5,
                        row = 1,
                        col = 2,
                        intensity = 6)

# Change to directory for output
setwd('/home/christinehicks/maanova/output')

# Save data to check it managed to read in correctly
save(man_data, file = './InputData.RData')

# Gridcheck, arrayview and RIPlot to check the quality
gridcheck(man_data)
riplot(man_data)
```

```

arrayview(man.data)

# Transform data to remove anomalies
transformed.data <- transform.madata(man.data,
                                     method = 'glowess',
                                     f=0.1,
                                     draw='pdf')
transformed.data <- transform.madata(transformed.data,
                                     method='rLOWESS',
                                     f=0.1,
                                     draw='pdf')

# Check the final result in graphical form
arrayview(transformed.data)
file.rename('./output/ArrayView',
            './output/ArrayView after Regional LOWESS')

# Save data ready for model fitting
save(transformed.data,
      file='./Ready_for_model_fitting.RData')

# Fit a model to the data based on the terms of variation given
anova <- fitmaanova(transformed.data,
                    ~Dye+Array+(Time*Treatment)/BioRep,
                    ~Dye+Array)

save(anova, './AfterAnova.RData')

print('MAANOVA done. Starting F-tests..')

# Do F-tests on the terms to find out how much variation they provide to the model
# Remove f-tests as they are complete because they take up a lot of RAM.
ftest4 <- matest(transformed.data,
                 anova,
                 'Time:Treatment:BioRep',
                 n.perm=1)
save(ftest4, file='./output/F-Test Time x Treatment x BioRep.RData')

ftest1 <- matest(transformed.data,
                 anova,
                 'Time',

```

```

        n.perm=1)
save(ftest1, file='./output/F-Test Time.RData')
test1 <- convertmatest(ftest4, ftest1)
rm(ftest1)

ftest2 <- matest(transformed.data,
                 anova,
                 'Treatment',
                 n.perm=1)
save(ftest2, file='./output/F-Test Treatment.RData')
test2 <- convertmatest(ftest4, ftest2)
rm(ftest2)

ftest3 <- matest(transformed.data, anova,
                 'Time:Treatment',
                 n.perm=1)
save(ftest3, file='./output/F-Test Time x Treatment.RData')
test3 <- convertmatest(ftest4, ftest3)
rm(ftest3, ftest4)

# Adjust P Values for false discovery rate
test1 <- adjPval(test1, 'stepdown')
test2 <- adjPval(test2, 'stepdown')
test3 <- adjPval(test3, 'stepdown')

# Test the terms and draw a Venn Diagram
analysematest('Fs', test1, test2, test3, useAdjPVals=T)

# Adjust p values for FDR
test1 <- adjPval(test1, 'stepdown')
test2 <- adjPval(test2, 'stepdown')
test3 <- adjPval(test3, 'stepdown')

# Test the terms and draw a venn diagram
analysematest('Fs', test1, test2, test3, useAdjPVals=T)

```

Appendix B

DESeq2 analysis script

This R script is the top-level script for the differential gene expression of RNAseq data described in Chapter 5.

```
library('DESeq2')

setwd('~/Documents/DESeq2')
data = read.table('~/Documents/DESeq2/ReadCounts.txt', header=T)
dir.create('Results')

#Create design
condition = factor(c(rep('control',3), rep('salt',3)))
type = rep('paired-end',6)
coldat = data.frame(condition,type)

#Apply differential expression testing
rownames(data) <- data$GeneID
data$GeneID <- NULL
row.names(coldat) = names(data)
cds = DESeqDataSetFromMatrix(countData = data,
                             colData=coldat, design = condition)
cds = DESeq(cds)
res = results(cds)

# Export results
write.table(res,paste('Results/DEGs.txt',sep='\t'))

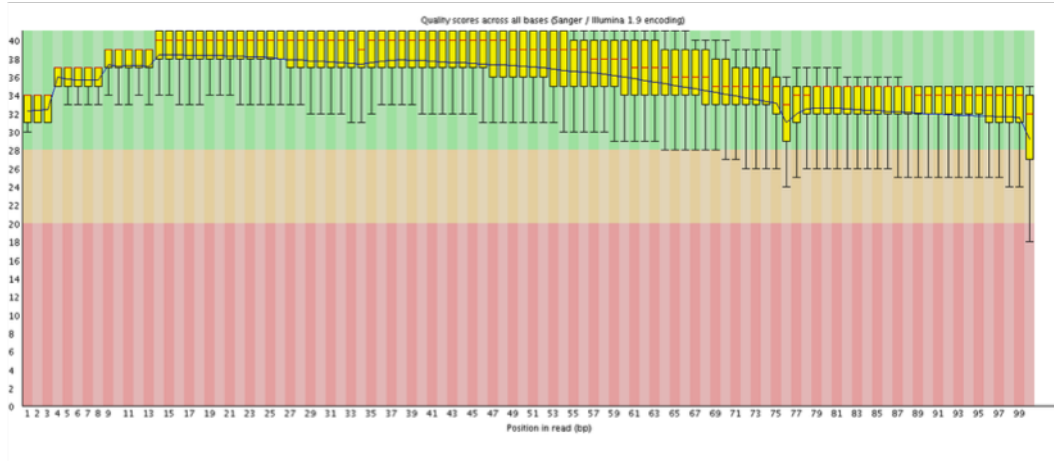
#Normalize counts
rld <- rlogTransformation(cds, blind=TRUE)
vsd <- varianceStabilizingTransformation(cds, blind=FALSE)
```

```
norm <- normTransform(cds)
log2.vds.counts <- assay(vsd)
log2.norm.counts <- assay(norm)

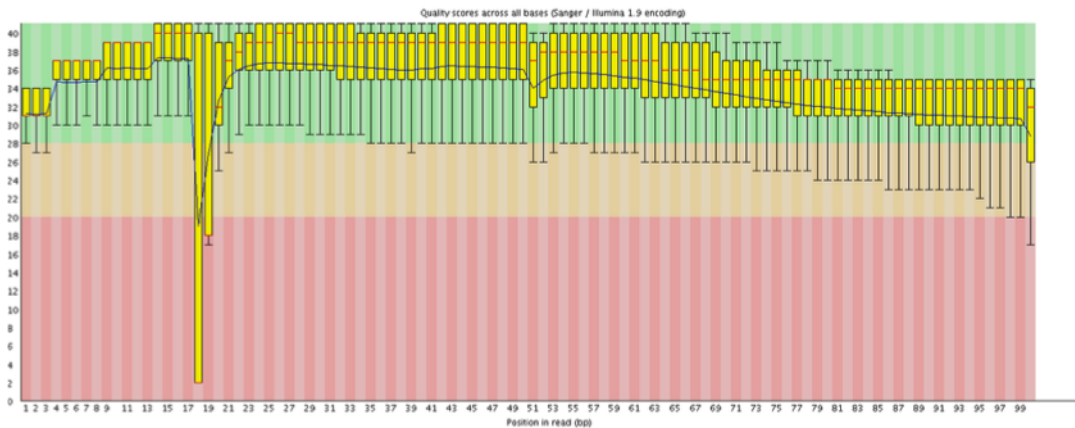
# Export normalized counts
write.table(log2.norm.counts, 'log2.norm.counts.txt', sep='\t')
write.table(log2.vsd.counts, 'log2.vsd.counts.txt', sep='\t')
```

Appendix C

Quality control of GD33DH RNAseq reads



(a) Per base quality assessment of R1 reads for sample S02 L01



(b) Per base quality assessment of R2 reads for sample S02 L01

Figure C.1: Quality assessment of RNAseq reads

Quality of the (a) forward (R1) reads and (b) reverse (R2) reads described in Chapter 3. Phred score on the y -axis, position of base pair in read along the x -axis. The colours indicate quality score, red for 'Good' ($>Q30$), orange is 'Average' ($Q20-28$ and red is 'Poor' ($<Q20$).

Appendix D

Transcript quality control of *de novo* assembly

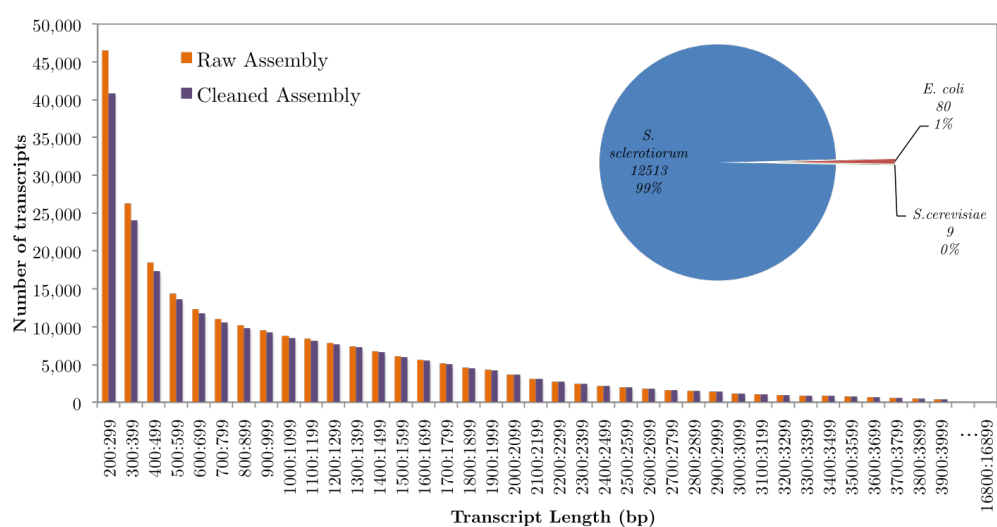


Figure D.1: Sequence distribution of *B. oleracea* GD33DH *de-novo* assembly and sources of contamination

The bar chart shows the number of transcripts found at 100 bp intervals in the raw *de-novo* assembly (orange bars) and the cleaned assembly (purple bars), described in Chapter 3. A break down of the origin of contaminating sequences is given the piechart, with percentage and count of transcripts not belonging to GD33DH in the assembly.

Appendix E

Differentially expressed transcription factors in time-series experiment

TF family	Up-regulated	Arabidopsis ortholog
AP2-EREBP	13	<i>RAP2.10</i> ; <i>ERF4</i> ; <i>ERF13</i>
ARF	6	<i>ARF6</i> ; <i>ARF10</i> ; <i>ARF16</i> ; <i>ARF18</i>
bHLH	25	<i>AIB</i> ; <i>BPEub</i> ; <i>HEC1</i> ; <i>ICE1</i> ; <i>LRL1</i> ; <i>LRL2</i> ; <i>PIL5</i> ; <i>MYC2</i>
bZIP	19	<i>ABF3</i> ; <i>ABF4</i> ; <i>ABI5</i> ; <i>AREB3</i> ; <i>AHBP-1B</i> ; <i>GBF3</i> ; <i>TGA1</i> ; <i>TGA3</i>
C2H2	9	<i>IDD4</i> ; <i>STZ</i> ; <i>ZF4</i>
CCAAT	16	<i>NF-YA2</i> ; <i>NF-YA5</i> ; <i>NF-YB1</i> ; <i>NF-YB8</i> ; <i>NF-YB10</i> ; <i>HTA13</i> ; <i>SNARE-like</i>
G2-like	17	<i>KAN</i>
HB	17	<i>BLH1</i> ; <i>BLH4</i> ; <i>HAT22</i> ; <i>HB-12</i> ; <i>HB6</i> ; <i>KNAT3</i> ; <i>KNAT4</i>
MADS	11	<i>ALG20</i> ; <i>FLC</i> ; <i>MAF3</i> ; <i>SEP4</i> ; <i>SVP</i>
MYB and MYB-related	27	<i>MYB3</i> ; <i>MYB32</i> ; <i>MYB34</i> ; <i>MYB46</i> ; <i>MYB47</i> ; <i>MYB55</i> ; <i>MYB96</i> ; <i>GYRB3</i> ; <i>TRB3</i>
NAC	24	<i>ATAF2</i> ; <i>NAC036</i> ; <i>NAC095</i> ; <i>NAC102</i>
Orphans	4	<i>DAR2</i> ; <i>RR16</i>
WRKY	9	<i>WRKY2</i> ; <i>WRK21</i> ; <i>WRKY28</i> ; <i>WRKY33</i> ; <i>WRKY41</i> ; <i>WRKY48</i>

Table E.1: Top up-regulated TF families

Number of up-regulated transcripts in selected TF families and named Arabidopsis orthologs where available. Genes in bold have previously been implicated in abiotic stress and have been discussed in the results sections of Chapter 4.

TF family	Down-regulated	Arabidopsis ortholog
AP2-EREBP	23	<i>TEM1</i> ; <i>TOE3</i> ; <i>ERF5</i> ; <i>TINY2</i> ; <i>DREB2B</i> ; <i>CRF4</i>
ARF	7	<i>ARF2</i> ; <i>ARF4</i> ; <i>MP</i>
bHLH	37	<i>PYE</i> ; <i>PIF4</i> ; <i>PIF7</i> ; <i>FMA</i> ; <i>LHW</i> ; <i>bHLH34</i> ; <i>SPT</i> ; <i>UNE10</i>
bZIP	23	<i>bZIP17</i> ; <i>bZIP23</i> ; <i>bZIP25</i> ; <i>bZIP49</i> ; <i>bZIP68</i> ; <i>TGA4</i> ; <i>GBF1</i>
C2H2	12	<i>IDD5</i> ; <i>IDD11</i> ; <i>TFIIIA</i> ; <i>SGR5</i>
CCAAT	9	<i>HTA13</i> ; <i>NF-YA2</i> ; <i>NF-YA5</i> ; <i>NF-YB1</i>
G2-like	11	<i>KAN3</i> ; <i>PHR1</i> ; <i>GLK2</i>
HB	27	<i>BLH7</i> ; <i>HAT9</i> ; <i>HB16</i> ; <i>HB18</i> ; <i>HB34</i> ; <i>HB5</i> ; <i>HDG2</i> ; <i>KNAT7</i> ; <i>PDF2</i> ; <i>RPL</i> ; <i>WOX1</i> ; <i>WUS</i>
MADS	8	-
MYB and MYB-related	50	<i>ADA2A</i> ; <i>MYB10</i> ; <i>MYB17</i> ; <i>MYB30</i> ; <i>MYB31</i> ; <i>MYB59</i> ; <i>MYB60</i> ; <i>MYB90</i> ; <i>MYB108</i> ; <i>MYBL2</i> ; <i>LHY</i> ; <i>RL4</i> ; <i>SWI3B</i>
NAC	15	<i>CUC1</i> ; <i>LOV1</i> ; <i>NAC028</i> ; <i>NAC050</i> ; <i>NAC096</i> ; <i>SDH2-2</i> ; <i>SOG1</i>
Orphans	23	<i>ARR4</i> ; <i>ARR9</i> ; <i>CIA2</i> ; <i>CIL</i>
WRKY	12	<i>WRKY3</i> ; <i>WRKY4</i> ; <i>WRKY15</i> ; <i>WRKY26</i> ; <i>WRKY32</i> ; <i>WRKY39</i> ; <i>WRKY69</i>

Table E.2: Top down-regulated TF families

Number of down-regulated transcripts in selected TF families and named Arabidopsis orthologs where available. Genes in bold have previously been implicated in abiotic stress and have been discussed in the results of Chapter 4.

Appendix F

Cluster plots

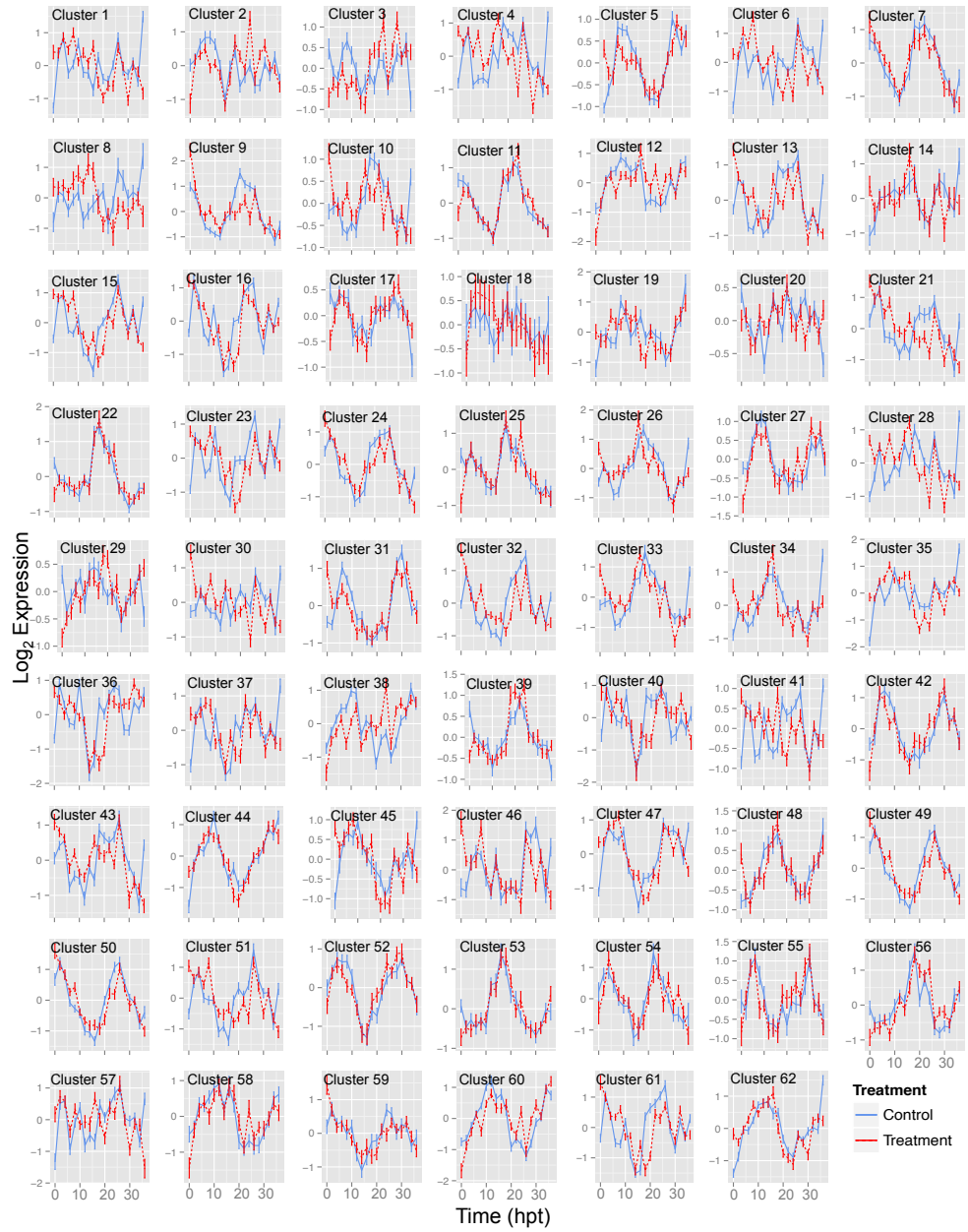


Figure F.1: Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in *B. oleracea* GD33DH

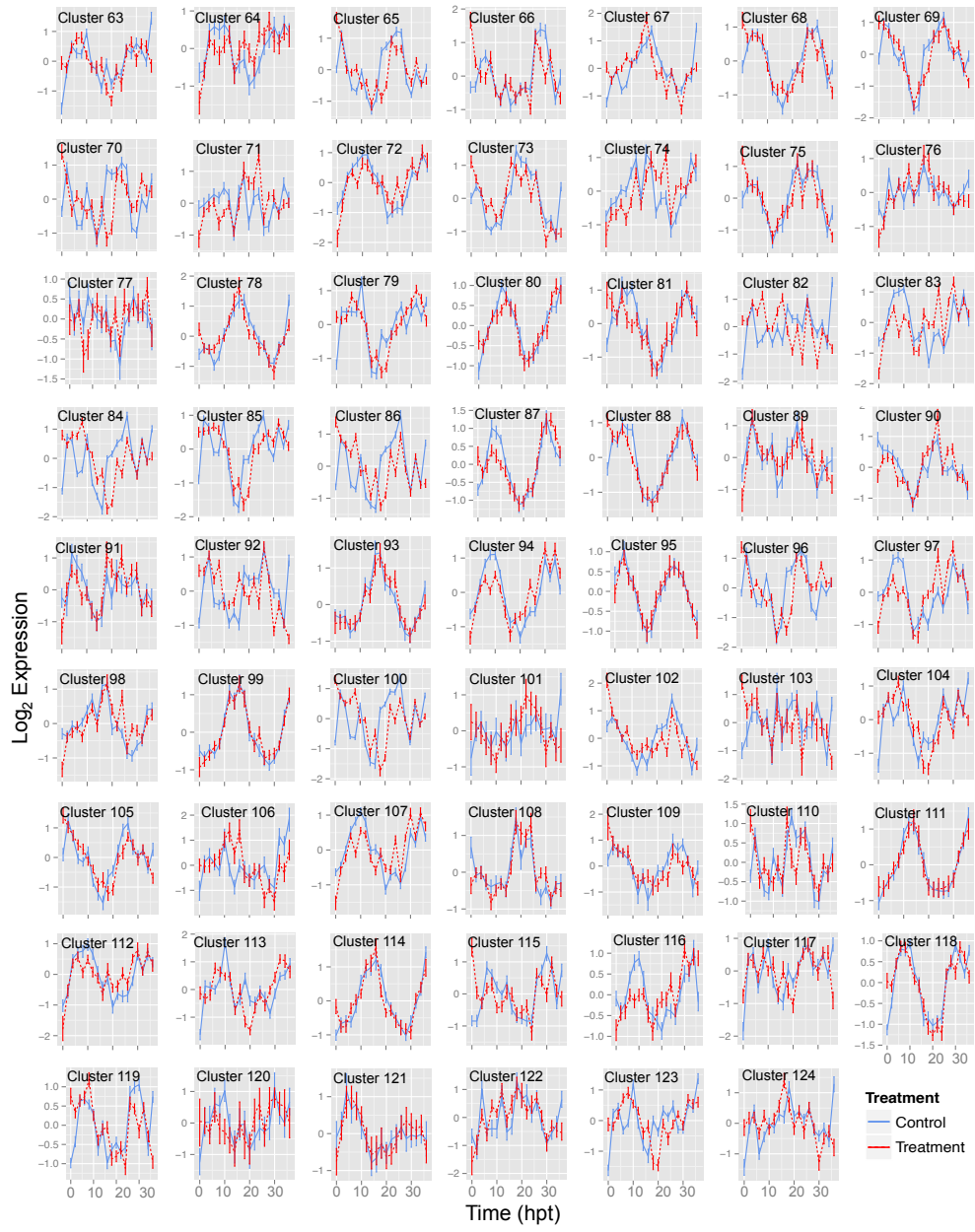


Figure F.1: Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in *B. oleracea* GD33DH

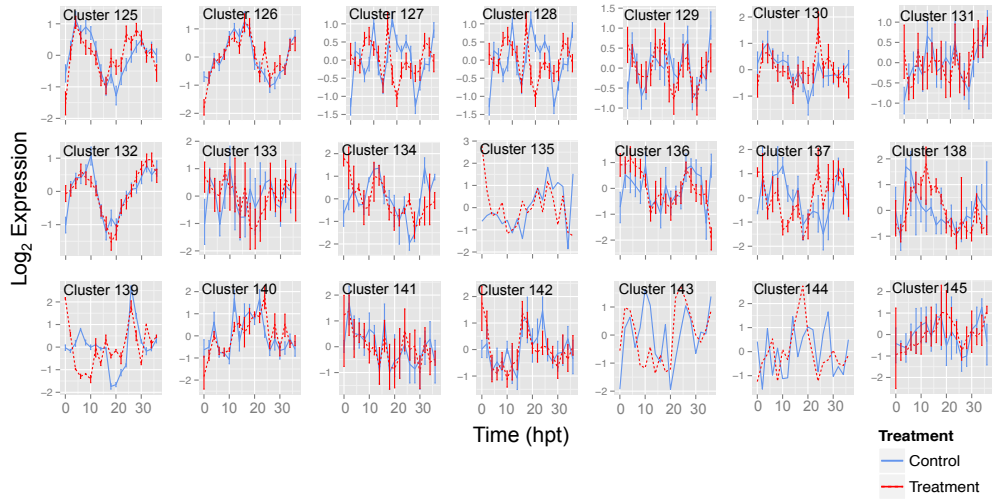


Figure F.1: Plots of the mean gene expression profile of co-expressed transcripts differentially expressed after salt shock in *B. oleracea* GD33DH

Both treatment and control expression data described in Chapter 4 was clustered alongside using the Multiple Data Integration algorithm. The red line represents the mean expression profile for the salt-treated transcripts and blue the control. The error bar is the 99% confidence interval of the data within each cluster. Note scale may be different for each plot. Data were mean centred and standard deviation normalised separately for each condition prior to clustering. The x -axis is hours post treatment (hpt) and the y -axis is the \log_2 expression.

Appendix G

GO terms associated with clusters

Cluster name	Total cluster	Top GO Term	Corrected p-value	Cluster frequency	Total frequency
100	241	Translation	0	70.3%	3.7%
84	161	Translation	2.5461E-58	62.5%	3.7%
61	133	Translation	2.6803E-49	58.8%	3.7%
86	220	Translation	5.9035E-24	34.8%	3.7%
6	134	IPP biosynthetic process	1.4297E-21	28.9%	1.2%
82	149	Photosynthesis	3.0147E-17	27.7%	2%
65	207	Translation	2.1682E-14	24.8%	3.7%
128	157	Translation	2.8897E-14	29%	3.7%
16	157	Nucleotide biosynthetic process	6.6687E-14	24.6%	2%
15	142	Cellular N compound biosynthetic process	3.5521E-13	39.7%	6.5%
140	12	Regulation of endopeptidase activity	6.1188E-13	75%	0.2%
96	112	Ribosome biogenesis	9.1836-12	26.1%	2%
92	160	Isoprenoid biosynthetic process	2.3595E-09	22.7%	2.7%
123	137	Translation	6.3124E-09	23.8%	3.7%
89	56	Regulation of endopeptidase activity	7.1148E-08	18.7%	0.2%
47	172	Histidine metabolic process	2.1649E-07	5.8%	0%
112	107	Autophagy	9.3787E-07	12.9%	0.3%
79	159	Proteasome assembly	1.8998E-06	13.4%	1.2%
36	137	RNA methylation	2.6381E-06	17.6%	1.8%
13	177	IPP biosynthetic process	3.4628E-06	12.9%	1.2%
73	129	Protein-chromophore linkage	9.2068E-06	6.5%	0.1%
99	122	Myo-inositol hexakisphosphate biosynthetic process	1.1473E-05	9.2%	0.2%
69	97	Riboflavin and derivative metabolic process	1.1993E-05	7.5%	0%
41	178	Photorespiration	1.2592E-05	11.3%	1.1%
14	50	Response to boron	1.4741E-05	14.2%	0%
75	60	Aromatic compound catabolic process	1.7611E-05	23%	0.7%
67	128	Generation of precursor metabolites and energy	3.069E-05	22.3%	3.8%
138	5	Translational termination	3.7814E-05	66.6%	0%
37	135	tRNA metabolic process	4.9803E-05	12.5%	1.1%
49	105	Defense response to Gram-negative bacterium	5.3383E-05	7.6%	0%
4	173	Photosynthesis, dark reaction	6.6618E-05	5.7%	0.1%
45	101	Epithelium development	7.7068E-05	4.7%	0%
113	92	Primary metabolic process	1.0092E-04	92.5%	61.7%
35	218	Generation of precursor metabolites and energy	1.1906E-04	16.2%	3.8%
106	22	Ion transmembrane transport	1.2655E-04	50%	0.9%
22	135	Response to pH	2.0427E-04	5%	0%
3	110	Inositol trisphosphate metabolic process	2.0907E-04	8%	0.1%

Cluster name	Total cluster	Top GO Term	Corrected p-value	Cluster frequency	Total frequency
85	184	Cellular N compound metabolic process	2.1348E-04	53.9%	28.1%
26	119	Small molecule metabolic process	05.297E-04	53.2%	23.9%
32	191	Translation	5.6582E-04	16%	3.7%
108	79	Response to aluminum ion	5.8576E-04	9%	0%
87	89	Epidermal cell differentiation	6.8081E-04	19.6%	3.1%
33	123	Regulation of photosynthesis	7.3364E-04	5.8%	0.1%
142	11	Proteolysis	8.9223E-04	100%	7.5%
110	88	Mucilage extrusion from seed coat	0.001376	5.5%	0%
119	142	Threonine catabolic process	0.0015117	3.9%	0%
117	97	Regulation of nitrate assimilation	0.0016307	4.6%	0%
76	120	Lipid transport	0.0017235	10.4%	0.8%
59	135	Phosphatidylethanolamine biosynthetic process	0.0017427	4.2%	0%
101	34	Cell wall cellulose biosynthetic process	0.0018383	12.5%	0%
43	126	Photosynthesis, light reaction	0.0020946	11.3%	1.4%
57	100	Regulation of nitrate assimilation	0.0024812	3.7%	0%
50	123	Cyanide catabolic process	0.0030483	3.1%	0%
23	198	Cellular macromolecule metabolic process	0.0030956	63.4%	41%
62	196	Alkane biosynthetic process	0.0032655	2.7%	0%
44	98	Glyphosate metabolic process	0.0048349	3.9%	0%
88	84	JA and ethylene-dependent systemic resistance	0.0048433	7.1%	0%
70	130	Embryonic development	0.0059028	17.4%	3.9%
24	123	Ribosome biogenesis	0.0083308	12.1%	2%
54	67	Lipid particle organization	0.0092652	8%	0%
107	251	Drug transmembrane transport	0.009825	5.3%	0.5%
133	7	Gibberellic acid mediated signaling pathway	0.010031	40%	0.5%
28	192	Stomatal complex morphogenesis	0.010653	7.3%	0.8%
132	79	Lipid particle organization	0.011821	4.5%	0%
141	6	Oligopeptide transport	0.013173	50%	0.7%
48	89	Mucilage metabolic process	0.013641	5.8%	0.1%
97	134	Establishment of cell polarity	0.013699	3.6%	0%
38	168	Fatty acid beta-oxidation	0.013888	7.7%	0.9%
83	194	Cellular catabolic process	0.014901	24.2%	9.5%
104	129	Cellular biosynthetic process	0.016156	58.4%	34.1%
78	102	Chlorophyll metabolic process	0.016464	11.3%	0.9%
129	36	Cell plate formation involved in plant-type cell wall biogenesis	0.018536	11.1%	0%
40	79	Nuclear export	0.018577	8.8%	0.4%
63	119	Phyllome development	0.022397	14.7%	3.1%

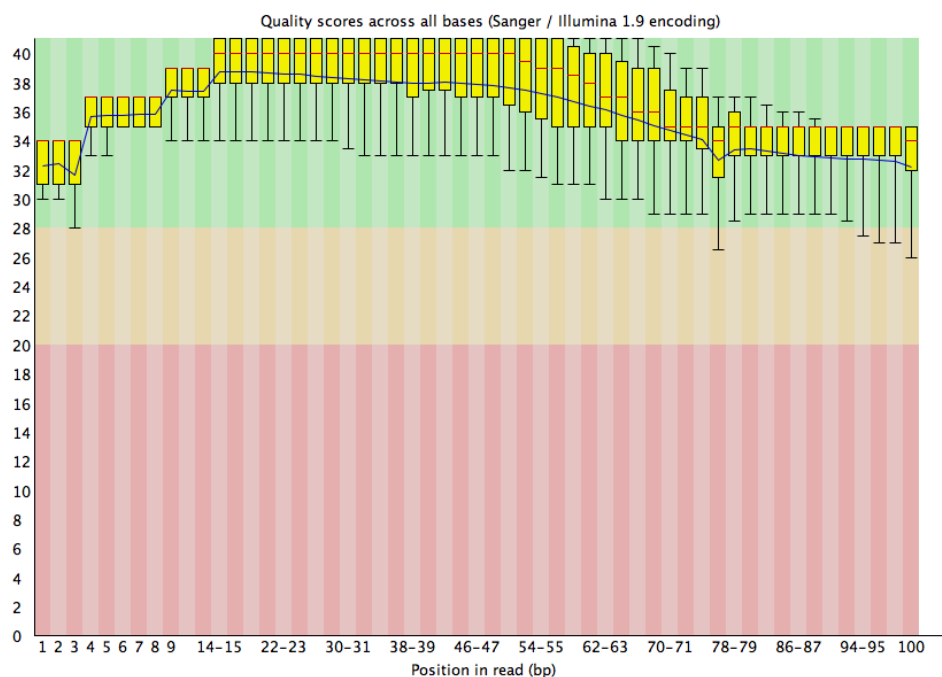
Cluster name	Total cluster	Top GO Term	Corrected p-value	Cluster frequency	Total frequency
105	77	RNA stabilization	0.02266	5%	0%
93	84	Microsporocyte differentiation	0.023137	4.2%	0%
27	120	Organic ether metabolic process	0.023581	7.1%	0.4%
51	107	Potassium ion import	0.025115	3.8%	0%
111	64	Riboflavin biosynthetic process	0.025338	4.8%	0%
68	120	JA acid and ethylene-dependent systemic resistance	0.02638	4.2%	0%
109	47	Regulation of DNA methylation	0.026814	8.3%	0%
30	158	Carotenoid catabolic process	0.028213	2.5%	0%
25	98	Cell death	0.028812	9.8%	0.8%
98	174	Abscisic acid mediated signaling pathway	0.028975	9.7%	1.7%
116	120	Glyoxysome organization	0.031207	3.8%	0%
34	135	Polysaccharide metabolic process	0.032945	18%	5.2%
115	102	Negative regulation of circadian rhythm	0.034552	4.1%	0%
80	133	Divalent metal ion transport	0.036747	10.9%	1.4%
121	38	Cysteine biosynthetic process from serine	0.039201	8.6%	0%
114	83	Response to cadmium ion	0.040837	18.3%	3.5%
81	48	AMP metabolic process	0.043333	6.6%	0%
11	140	Glucosinolate catabolic process	0.044652	4.2%	0.1%
7	57	Isoprenoid biosynthetic process	0.045202	13.7%	1.2%
55	110	Peroxisome fission	0.04524	3.4%	0%
74	97	Nickel ion transport	0.048202	4.6%	0%

Table G.1: Over-represented GO terms of clusters

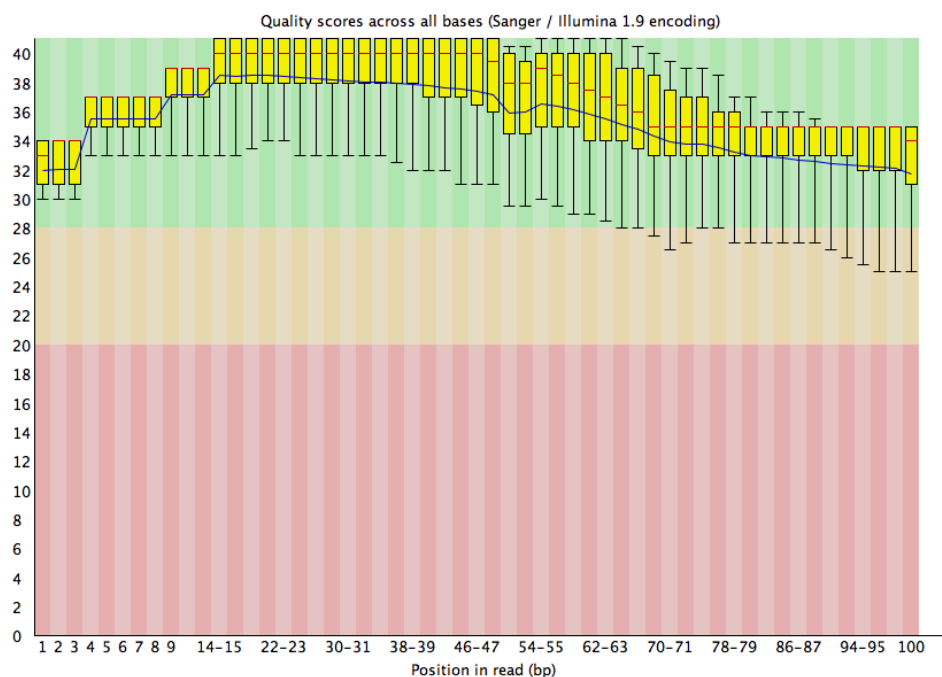
Significantly over-represented GO terms for each cluster produced using MDI, as described in Chapter 4, in ascending order of significance.

Appendix H

Quality control of wild *Brassica* species RNAseq reads



(a) Per base quality assessment of R1 reads for C13001 salt-treated (Rep A)



(b) Per base quality assessment of R2 reads for C13001 salt-treated (Rep A)

Figure H.1: Quality assessment of RNAseq reads

An exemplar quality plot of (a) forward (R1) reads and (b) reverse (R2) reads described in Chapter 5. Phred score on the y -axis, position of base pair in read along the x -axis. The colours indicate quality score, red for 'Good' ($>Q30$), orange is 'Average' ($Q20-28$ and red is 'Poor' ($<Q20$).

Appendix I

Additional data files

Included on the CD are data files relating to:

1. A multi-fasta file containing the transcripts used in the microarray design.
2. The TO1000 GO annotation for identifying over-represented GO terms using BiNGO.
3. Differentially expressed transcripts identified from time-series experiment (Chapter 4).
4. Differentially expressed genes identified from RNAseq of wild C-genome *Brassica* species (Chapter 5).

Bibliography

- AbdElgawad, H., Zinta, G., Hegab, M. M., Pandey, R., Asard, H. and Abuelsoud, W. (2016). High Salinity Induces Different Oxidative Stress and Antioxidant Responses in Maize Seedlings Organs. *Frontiers in Plant Science* **7**, 276.
- Abe, H., Takeshi, U., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscissic Acid Signaling. *The Plant Cell* **15**, 63–78.
- Abeles, A. B., Morgan, P. W. and Saltveit, M. E. J. (1992). *Ethylene in Plant Biology*. San Diego, CA: Academic Press.
- Abreu, E. F. M. and Aragao, F. J. L. (2007). Isolation and Characterization of a myo-inositol-1-phosphate Synthase Gene from Yellow Passion Fruit (*Passiflora edulis* f. *flavicarpa*) Expressed During Seed Development and Environmental Stress. *Annals of Botany* **99**, 285–292.
- Adam, K. L., Cronn, R. C., Percifield, R. and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100**, 4649–4654.
- Ahuja, I., de Vos, R. C. H., Bones, A. M. and Hall, R. D. (2010). Plant molecular stress responses face climate change. *Trends in Plant Science* **15**, 664–674.
- Akama, S., Shimizu-Inatsugi, R., Shimizu, K. K. and Sese, J. (2014). Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. *Nucleic Acids Research* **42**, e46–e46.
- Akter, N., Islam, M. R., Karim, M. A. and Hossain, T. (2014). Alleviation of drought stress in maize by exogenous application of gibberellic acid and cytokinin. *Journal of Crop Science and Biotechnology* **17**, 41–48.
- Ali, Y., Aslam, Z., Ashraf, M. Y. and Tahir, G. R. (2004). Effect of salinity on chlorophyll concentration, leaf area, yield and yield components of rice genotypes grown under saline environment. *International Journal of Environmental Science & Technology* **1**, 221–225.
- Allakhverdiev, S. I. and Murata, N. (2004). Environmental stress inhibits the synthesis de novo of proteins involved in the photodamage–repair cycle of Photosystem II in *Synechocystis* sp. PCC 6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1657**, 23–32.
- Allakhverdiev, S. I., Nishiyama, Y., Miyairi, S., Yamamoto, H., Inagaki, N., Kaneshaki, Y. and Murata, N. (2002). Salt Stress Inhibits the Repair of Photodamaged Photosystem II by

- Suppressing the Transcription and Translation of psbA Genes in *Synechocystis*. *Plant Physiology* **130**, 1443–1453.
- Allen, G. J., Chu, S. P., Schumacher, K., Shimazaki, C. T., Vafeados, D., Kemper, A., Hawke, S. D., Tallman, G., Tsien, R. Y., Harper, J. F. et al. (2000). Alteration of Stimulus-Specific Guard Cell Calcium Oscillations and Stomatal Closing in *Arabidopsis det3* Mutant. *Science* **289**, 2338–2342.
- An, F., Zhao, Q., Ji, Y., Li, W., Jiang, Z., Yu, X., Zhang, C., Han, Y., He, W., Liu, Y. et al. (2010). Ethylene-Induced Stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 Is Mediated by Proteasomal Degradation of EIN3 Binding F-Box 1 and 2 That Requires EIN2 in *Arabidopsis*. *The Plant Cell* **22**, 2384–2401.
- Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* **11**, R106.
- Anderson, J. P., Badrunzaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., Maclean, D. J., Ebert, P. R. and Kazan, K. (2004). Antagonistic Interaction between Abscissic Acid and Jasmonate-Ethylene Signaling Pathways Modulates Defense Gene Expression and Disease Resistance in *Arabidopsis*. *The Plant Cell* **16**, 3460–3479.
- Andriolo, J. L., da Luz, G. L., Witter, M. H., Godoi, R. S., Barros, G. T. and Bortolotto, O. C. (2005). Growth and yield of lettuce plants under salinity. *Horticultura Brasileira* **23**, 931–934.
- AQUASTAT (2014). Infographics on Water Resources and Uses.
- Arc, E., Sechet, J., Corbineau, F., Rajjou, L. and Marion-Poll, A. (2013). ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. *Frontiers in Plant Science* **4**, 63.
- Arfan, M. (2009). Exogenous Application of Salicylic Acid through Rooting Medium Modulates Ion Accumulation and Antioxidant Activity in Spring Wheat under Salt Stress. *International Journal of Agriculture & Biology* **11**, 437–442.
- Ashburner, M. and Lewis, S. (2002). On ontologies for biologists: the Gene Ontology - untangling the web. *‘In Silico’ Simulation of Biological Processes: Novartis Foundation Symposium* **247**, 66–83.
- Ashraf, M. and Harris, P. J. C. (2013). Photosynthesis under stressful environments: An overview. *Photosynthetica* **51**, 163–190.
- Atkinson, L. D., Hilton, H. W. and Pink, D. A. C. (2012). A study of variation in the tendency for postharvest discoloration in a lettuce (*Lactuca sativa*) diversity set. *International Journal of Food Science & Technology* **48**, 801–807.
- Atkinson, N. J. and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany* **63**, 3523–3543.
- Attia, H., Arnaud, N., Karray, N. and Lachaâl, M. (2008). Long-term effects of mild salt stress on growth, ion accumulation and superoxide dismutase expression of *Arabidopsis* rosette leaves. *Physiologia Plantarum* **132**, 293–305.

- Attia, H., Karray, N., Msilini, N. and Lachaâl, M.** (2011). Effect of salt stress on gene expression of superoxide dismutases and copper chaperone in *Arabidopsis thaliana*. *Biologia Plantarum* **55**, 159–163.
- Babitha, K. C., Ramu, S. V., Pruthvi, V., Mahesh, P., Nataraja, K. N. and Udayakumar, M.** (2012). Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in *Arabidopsis*. *Transgenic Research* **22**, 327–341.
- Balazadeh, S., Siddiqui, H., Allu, A. D., Matallana-Ramirez, L. P., Caldana, C., Mehrnia, M., Zanol, M. I., Köhler, B. and Mueller-Roeber, B.** (2010). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal* **62**, 250–264.
- Baldoni, E., Genga, A. and Cominelli, E.** (2015). Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. *International Journal of Molecular Sciences* **16**, 15811–15851.
- Barah, P., Mahantesha, N., Jayavelu, N. D., Sowdhamini, R., Shameer, K. and Bones, A. M.** (2016). Transcriptional regulatory networks in *Arabidopsis thaliana* during single and combined stresses. *Nucleic Acids Research* **44**, 3147–3164.
- Barker, G. C., Larson, T. R., Graham, I. A., Lynn, J. R. and King, G. J.** (2007). Novel Insights into Seed Fatty Acid Synthesis and Modification Pathways from Genetic Diversity and Quantitative Trait Loci Analysis of the Brassica C Genome. *Plant Physiology* **144**, 1827–1842.
- Barker, M. S., Vogel, H. and Schranz, M. E.** (2009). Paleopolyploidy in the Brassicales: Analyses of the Cleome Transcriptome Elucidate the History of Genome Duplications in *Arabidopsis* and Other Brassicales. *Genome Biology and Evolution* **1**, 391–399.
- Barragan, V., Leidi, E. O., Andres, Z., Rubio, L., De Luca, A., Fernandez, J. A., Cubero, B. and Pardo, J. M.** (2012). Ion Exchangers NHX1 and NHX2 Mediate Active Potassium Uptake into Vacuoles to Regulate Cell Turgor and Stomatal Function in *Arabidopsis*. *The Plant Cell* **24**, 1127–1142.
- Barrero, J. M., Rodrigues, P. L., Quesada, V., Piqueras, P., Ponce, M. R. and Micol, J. L.** (2006). Both abscisic acid (ABA)-dependent and ABA-independent pathways govern the induction of NCED3, AAO3 and ABA1 in response to salt stress. *Plant, Cell and Environment* **29**, 2000–2008.
- Bassil, E., Coku, A. and Blumwald, E.** (2012). Cellular ion homeostasis: emerging roles of intracellular NHX Na⁺/H⁺ antiporters in plant growth and development. *Journal of Experimental Biology* **63**, 5727–5740.
- Bassil, E., Tajima, H., Liang, Y. C., Ohto, M. a., Ushijima, K., Nakano, R., Esumi, T., Coku, A., Belmonte, M. and Blumwald, E.** (2011). The *Arabidopsis* Na⁺/H⁺ Antiporters NHX1 and NHX2 Control Vacuolar pH and K⁺ Homeostasis to Regulate Growth, Flower Development, and Reproduction. *The Plant Cell* **23**, 3482–3497.
- Batelli, G., Verslues, P. E., Agius, F., Qiu, Q., Fujii, H., Pan, S., Schumaker, K. S., Grillo, S. and Zhu, J. K.** (2007). SOS2 Promotes Salt Tolerance in Part by Interacting with the Vacuolar H⁺-ATPase and Upregulating Its Transport Activity. *Molecular and Cellular Biology* **27**, 7781–7790.

- Bechtold, U., Penfold, C. A., Jenkins, D. J., Legaie, R., Moore, J. D., Lawson, T., Matthews, J. S. A., Vialet-Chabrand, S. R. M., Baxter, L., Subramaniam, S. et al. (2016). Time-Series Transcriptomics Reveals That AGAMOUS-LIKE22 Affects Primary Metabolism and Developmental Processes in Drought-Stressed Arabidopsis. *The Plant Cell* **28**, 345–366.
- Benito, B., Haro, R., Amtmann, A., Cuin, T. A. and Dreyer, I. (2014). Journal of Plant Physiology. *Journal of Plant Physiology* **171**, 723–731.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300.
- Benjamini, Y. and Hochberg, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics* **25**, 60–83.
- Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference* **82**, 163–170.
- Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J. and Parcy, F. (2002). The Homologous ABI5 and EEL Transcription Factors Function Antagonistically to Fine-Tune Gene Expression during Late Embryogenesis. *The Plant Cell* **14**, 1391–1403.
- Bernstein, L., Francois, L. E. and Clark, R. A. (1974). Interactive Effects of Salinity and Fertility on Yields of Grains and Vegetables. *Agronomy Journal* **66**, 412–421.
- Bhaskara, G. B., Nguyen, T. T. and Verslues, P. E. (2012). Unique Drought Resistance Functions of the Highly ABA-Induced Clade A Protein Phosphatase 2Cs. *Plant Physiology* **160**, 379–395.
- Binder, B. M., Walker, J. M., Gagne, J. M., Emborg, T. J., Hemmann, G., Bleecker, A. B. and Vierstra, R. D. (2007). The Arabidopsis EIN3 Binding F-Box Proteins EBF1 and EBF2 Have Distinct but Overlapping Roles in Ethylene Signaling. *The Plant Cell* **19**, 509–523.
- Birchler, J. A. and Veitia, R. A. (2007). The Gene Balance Hypothesis: From Classical Genetics to Modern Genomics. *The Plant Cell* **19**, 395–402.
- Birchler, J. A. and Veitia, R. A. (2012). Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences* **109**, 14746–14753.
- Blanc, G. and Wolfe, K. H. (2004). Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *The Plant Cell* **16**, 1679–1691.
- Blanco, G., Gerlagh, R., Suh, S., Barrett, J., de Coninck, H. C., Diaz Morejon, C. F., Mathur, R., Nakicenovic, N., Ofosu Ahenkora, A., Pan, J. et al. (2014). Drivers, Trends and Mitigation. In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1–62. Cambridge, UK: Cambridge University Press.
- Bleidorn, C. (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* **14**, 1–8.

- Bohuon, E., Keith, D. J., Parkin, I., Sharpe, A. G. and Lydiate, D. J.** (1996). Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. *Theoretical and Applied Genetics* **93**, 833–839.
- Bohuon, E., Ramsay, L. D., Craft, J. A., Arthur, A. E., Marshall, D. F., Lydiate, D. J. and Kearsey, M. J.** (1998). The Association of Flowering Time Quantitative Trait Loci with Duplicated Regions and Candidate Loci in *Brassica oleracea*. *Genetics* **150**, 393–401.
- Borkotoky, S., Saravanan, V., Jaiswal, A., Das, B., Selvaraj, S., Murali, A. and Lakshmi, P. T. V.** (2013). The Arabidopsis Stress Responsive Gene Database. *International Journal of Plant Genomics* **2013**, 949564.
- Bortesi, L. and Fischer, R.** (2015). Biotechnology Advances. *Biotechnology Advances* **33**, 41–52.
- Boursiac, Y., Chen, S., Luu, D. T., Sorieul, M., van den Dries, N. and Maurel, C.** (2005). Early Effects of Salinity on Water Transport in Arabidopsis Roots. Molecular and Cellular Features of Aquaporin Expression. *Plant Physiology* **139**, 790–805.
- Boutrot, F., Chantret, N. and Gautier, M. F.** (2008). Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. *BMC Genomics* **9**, 86.
- Bowers, J. E., Chapman, B. A., Rong, J. K. and Paterson, A. H.** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438.
- Brady, S. M., Sarkar, S. F., Bonetta, D. and McCourt, P.** (2003). The ABSCISIC ACID INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis. *The Plant Journal* **34**, 167–175.
- Bray, E. A.** (1993). Molecular Responses to Water Deficit. *Plant Physiology* **103**, 1035–1040.
- Breeze, E., Harrison, E., McHattie, S., Hughes, L., Hickman, R., Hill, C., Kiddle, S., Kim, Y. S., Penfold, C. A., Jenkins, D. et al.** (2011). High-Resolution Temporal Profiling of Transcripts during Arabidopsis Leaf Senescence Reveals a Distinct Chronology of Processes and Regulation. *The Plant Cell* **23**, 873–894.
- Broadley, M. R., Hammond, J. P., King, G. J., Astley, D., Bowen, H. C., Meacham, M. C., Mead, A., Pink, D. A. C., Teakle, G. R., Hayden, R. M. et al.** (2008). Shoot Calcium and Magnesium Concentrations Differ between Subtaxa, Are Highly Heritable, and Associate with Potentially Pleiotropic Loci in *Brassica oleracea*. *Plant Physiology* **146**, 1707–1720.
- Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P. O., Nam, H. G., Lin, J. F., Wu, S. H., Swidzinski, J., Ishizaki, K. et al.** (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. *The Plant Journal* **42**, 567–585.
- Burssens, S., Himanen, K., van de Cotte, B., Beeckman, T., Van Montagu, M., Inzé, D. and Verbruggen, N.** (2000). Expression of cell cycle regulatory genes and morphological alterations in response to salt stress in Arabidopsis thaliana. *Planta* **211**, 632–640.

- Cabot, C., Gallego, B., Martos, S., Barceló, J. and Poschenrieder, C.** (2013). Signal cross talk in Arabidopsis exposed to cadmium, silicon, and Botrytis cinerea. *Planta* **237**, 337–349.
- Calderhead, B. and Girolami, M.** (2009). Estimating Bayes factor via thermodynamic integration and population MCMC. *Computational Statistic & Data Analysis* **53**, 4028–4045.
- Carvalho, R. F., Campos, M. L. and Azevedo, R. A.** (2011). The Role of Phytochrome in Stress Tolerance. *J Integrative Plant Biology* **53**, 920–929.
- Cellier, F., Conejero, G., Ricaud, L., Luu, D. T., Lepetit, M., Gosti, F. and Casse, F.** (2004). Characterization of AtCHX17, a member of the cation/H⁺-exchangers, CHX family, from Arabidopsis thaliana suggests a role in K⁺-homeostasis. *The Plant Journal* **39**, 834–846.
- Černý, M., Jedelsky, P. L., Novak, J., Schlosser, A. and Brzobohaty, B.** (2014). Cytokinin modulates proteomic, transcriptomic and growth responses to temperature shocks in Arabidopsis. *Plant, Cell and Environment* **37**, 1641–1655.
- Chalhoub, B., Denoeud, F., Denoeud, F., Liu, S., Liu, S., Parkin, I. A. P., Parkin, I. A. P., Tang, H., Tang, H., Wang, X. et al.** (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**, 950–953.
- Chan, Z.** (2012). Expression profiling of ABA pathway transcripts indicates crosstalk between abiotic and biotic stress responses in Arabidopsis. *Genomics* **100**, 110–115.
- Chanroj, S., Lu, Y., Padmanaban, S., Nanatani, K., Uozumi, N., Rao, R. and Sze, H.** (2011). Plant-Specific Cation/H⁺ Exchanger 17 and Its Homologs Are Endomembrane K⁺ Transporters with Roles in Protein Sorting. *Journal of Biological Chemistry* **286**, 33931–33941.
- Chaves, M. M., Flexas, J. and Pinheiro, C.** (2008). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany* **103**, 551–560.
- Chen, H., Lai, Z., Shi, J., Xiao, Y., Chen, Z. and Xu, X.** (2010). Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant response to abscisic acid and abiotic stress. *BMC Plant Biology* **10**, 281.
- Chen, L., Song, Y., Li, S., Zhang, L., Zou, C. and Yu, D.** (2012). The role of WRKY transcription factors in plant abiotic stresses. *BBA - Gene Regulatory Mechanisms* **1819**, 120–128.
- Chen, X., Ge, X., Wang, J., Tan, C., King, G. J. and Liu, K.** (2015). Genome-wide DNA methylation profiling by modified reduced representation bisulfite sequencing in Brassica rapa suggests that epigenetic modifications play a key role in polyploid genome evolution. *Frontiers in Plant Science* **6**, 836.
- Chen, Z. and Gallie, D. R.** (2006). Dehydroascorbate Reductase Affects Leaf Growth, Development, and Function. *Plant Physiology* **142**, 775–787.
- Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G. and Wang, X.** (2012a). Biased Gene Fractionation and Dominant Gene Expression among the Subgenomes of Brassica rapa. *PLoS ONE* **7**, e36442.

- Cheng, F., Wu, J., Fang, L. and Wang, X.** (2012b). Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Frontiers in Plant Science* **3**, 198.
- Cheng, F., Wu, J., Schnable, J., Woodhouse, M. R., Liang, J., Cai, C., Freeling, M. and Wang, X.** (2016). Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytologist* **211**, 288–299.
- Cheng, F., Wu, J. and Wang, X.** (2014). Genome triplication drove the diversification of Brassica plants. *Horticulture Research* **1**, 14024.
- Cheng, N. H., Pittman, J. K., Barkla, B. J., Shigaki, T. and Hirschi, K. D.** (2003). The Arabidopsis *cax1* Mutant Exhibits Impaired Ion Homeostasis, Development, and Hormonal Responses and Reveals Interplay among Vacuolar Transporters. *The Plant Cell* **15**, 347–364.
- Cheng, N. H., Pittman, J. K., Shigaki, T., Lachmansingh, J., LeClere, S., Lahner, B., Salt, D. E. and Hirschi, K. D.** (2005). Functional Association of Arabidopsis CAX1 and CAX3 Is Required for Normal Growth and Ion Homeostasis. *Plant Physiology* **138**, 2048–2060.
- Chini, A., Fonseca, S., Chico, J. M., Fernández-Calvo, P. and Solano, R.** (2009). The ZIM domain mediates homo- and heteromeric interactions between Arabidopsis JAZ proteins. *The Plant Journal* **59**, 77–87.
- Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F. M., Ponce, M. R. et al.** (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666–671.
- Choi, W. G., Toyota, M., Kim, S. H., Hilleary, R. and Gilroy, S.** (2014). Salt stress-induced Ca²⁺ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proceedings of the National Academy of Sciences* **111**, 6497–6502.
- Churchill, G. A.** (2004). Using Anova to Analyze Microarray Data. *Biotechniques* **327**, 173–177.
- Comai, L.** (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**, 836–846.
- Combes, M. C., Dereeper, A., Severac, D., Bertrand, B. and Lashermes, P.** (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytologist* **200**, 251–260.
- Conant, G. C., Birchler, J. A. and Pires, J. C.** (2014). ScienceDirect Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology* **19**, 91–98.
- Conesa, A. and Götz, S.** (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics* **2008**, 619832.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X. et al.** (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**, 13.

- Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A. and Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology* **9**, R130.
- Cui, F., Brosche, M., Sipari, N., Tang, S. and Overmyer, K. (2013). Regulation of ABA dependent wound induced spreading cell death by MYB108. *New Phytologist* **200**, 634–640.
- Cusack, B. P. and Wolfe, K. H. (2006). Not Born Equal: Increased Rate Asymmetry in Relocated and Retrotransposed Rodent Gene Duplicates. *Molecular Biology and Evolution* **24**, 679–686.
- Das, K. and Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Frontiers in Environmental Science* **2**, 53.
- Daszkowska-Golec, A. and Szarejko, I. (2013). Open or close the gate – stomata action under the control of phytohormones in drought stress conditions. *Frontiers in Plant Science* **4**, 138.
- Davey, M. W., Graham, N. S., Vanholme, B., Swennen, R., May, S. T. and Keulemans, J. (2009). Heterologous oligonucleotide microarrays for transcriptomics in a non-model species; a proof-of-concept study of drought stress in Musa. *BMC Genomics* **10**, 436.
- de Abreu, C. E. B., Araújo, G. S., Monteiro-Moreira, A. C. O., Costa, J. H., Leite, H. B., Moreno, F. B. M. B., Prisco, J. T. and Gomes-Filho, E. (2014). Proteomic analysis of salt stress and recovery in leaves of *Vigna unguiculata* cultivars differing in salt tolerance. *Plant Cell Reports* **33**, 1289–1306.
- De Smet, I., Signora, L., Beeckman, T., Inze, D., Foyer, C. H. and Zhang, H. (2003). An abscisic acid-sensitive checkpoint in lateral root development of Arabidopsis. *The Plant Journal* **33**, 543–555.
- De Smet, R., Adam, K. L., Vandepoele, K., Van Montagu, M. C. E., Maere, S. and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences* **110**, 2898–2903.
- Delessert, C., Kazan, K., Wilson, I. W., van der Straeten, D., Manners, J., Dennis, E. S. and Dolferus, R. (2005). The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. *The Plant Journal* **43**, 745–757.
- Delmas, F., Sankaramarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J. G. B., McCourt, P. and Samuel, M. A. (2013). ABI3 controls embryo degreening through Mendel's *I* locus. *PNAS* **110**, E3888–E3894.
- Demidchik, V. and Maathuis, F. J. M. (2007). Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development. *New Phytologist* **175**, 387–404.
- Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V. and Toll, J. (2014). Adapting Agriculture to Climate Change: A Global Initiative to Collect, Conserve, and Use Crop Wild Relatives. *Agroecology and Sustainable Food Systems* **38**, 369–377.
- Deng-Hui, X., Zi-Bing, L., Zheng, Z. Y., M, V. K., Bao-Fang, F. and Zhi-Xiang, C. (2008). Stress- and Pathogen-Induced Arabidopsis WRKY48 is a Transcriptional Activator that Represses Plant Basal Defense. *Molecular Plant* **1**, 459–470.

- Dharmasiri, N., Dharmasiri, S. and Estelle, M.** (2005). The F-box protein TIR1 is an auxin receptor. *Nature* **435**, 441–445.
- Dietrich, P., Sanders, D. and Hedrich, R.** (2001). The role of ion channels in light-dependent stomatal opening. *Journal of Experimental Biology* **52**, 1959–1967.
- Dietz, K. J., Vogel, M. O. and Viehhauser, A.** (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma* **245**, 3–14.
- Ding, J. D., Yan, J. Y., Li, C. X., Li, G. X., Wu, Y. R. and Zheng, S. J.** (2015). Transcription factor WRKY46 modulates the development of Arabidopsis lateral roots in osmotic/salt stress conditions via regulation of ABA signaling and auxin homeostasis. *The Plant Journal* **84**, 56–69.
- Ding, Z. J., Yan, J. Y., Xu, X. Y., Yu, D. Q., Li, G. X., Zhang, S. Q. and Zheng, S. J.** (2014). Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in Arabidopsis. *The Plant Journal* **79**, 13–27.
- Donaldson, R. P., Kwak, Y., Yanik, T. and Sharma, V.** (2001). *Plant Peroxisomes and Glyoxysomes*. Encyclopedia of Life Science (ELS). Chichester, UK: John Wiley & Sons, Ltd.
- Downton, W. J. S. and Millhouse, J.** (1983). Turgor Maintenance during Salt Stress Prevents Loss of Variable Fluorescence in Grapevine Leaves. *Plant Science Letters* **31**, 1–7.
- Dugas, D. V., Monaco, M. K., Olsen, A., Klien, R. R., Kumari, S., Ware, D. and Klien, P. E.** (2011). Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid. *BMC Genomics* **12**, 514.
- Dunlap, J. R. and Binzel, M. L.** (1996). NaCl Reduces Indole-3-Acetic Acid Levels in the Roots of Tomato Plants Independent of Stress-Induced Absciscic Acid. *Plant Physiology* **112**, 379–384.
- Evans, A. R., Hall, D., Pritchard, J. and Newbury, H. J.** (2012). The roles of the cation transporters CHX21 and CHX23 in the development of Arabidopsis thaliana. *Journal of Experimental Botany* **63**, 59–67.
- Evans, M. J., Choi, W. G., Gilroy, S. and Morris, R. J.** (2016). A ROS-assisted Calcium Wave Dependent on AtRBOHD and TPC1 Propagates the Systemic Response to Salt Stress in Arabidopsis Roots. *Plant Physiology* **171**, 1771–1784.
- Faiyue, B., Vijayalaskshmi, C., Nawaz, S., Nagato, Y., Taketa, S., Ichii, M., Al-Azzawi, M. J. and Flowers, T. J.** (2010). Studies on sodium bypass flow in lateral rootless mutants lrt1 and lrt2, and crown rootless mutant crl1 of rice (Oryza sativa L.). *Plant, Cell and Environment* **33**, 687–701.
- FAO** (2003). Agriculture drainage water management in arid and semi-arid areas. *FAO Irrigation and Drainage Paper 61 Annex 1* pp. 1–13.
- FAO, IFAD and WFP** (2015). The State of Food Insecurity in the World 2015. Meeting the 2015 international hunger targets: taking stock of uneven progress. Technical report, Rome.
- Fenart, S., Chabi, M., Gallina, S., Huis, R., Neutelings, G., Riviere, N., Thomasset, B., Hawkins, S. and Lacau-Danila, A.** (2013). Intra-platform comparison of 25-mer and 60-mer oligonucleotide Nimblegen DNA microarrays. *BMC Research Notes* **6**, 43.

- Finkelstein, R. R., Tenbarger, K. M., Shumway, J. E. and Crouch, M. L. (1985). Role of ABA in Maturation of Rapeseed. *Plant Physiology* **78**, 630–636.
- Fischer, I., Steige, K. A., Stephan, W. and Mboup, M. (2013). Sequence Evolution and Expression Regulation of Stress-Responsive Genes in Natural Populations of Wild Tomato. *PLoS ONE* **8**, e78182.
- Fischer, R. A. (1968). Stomatal Opening: Role of Potassium Uptake by Guard Cells. *Science* **160**, 784–785.
- Flowers, T. J., Garcia, A., Koyama, M. and Yeo, A. R. (1997). Breeding for salt tolerance in crop plants - the role of molecular biology. *Acta Physiologiae Plantarum* **19**, 427–433.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**, 1531–1545.
- Forde, B. G. and Roberts, M. R. (2014). Glutamate receptor-like channels in plants: a role as amino acid sensors in plant defence? *F1000Prime Reports* **6**, 37.
- Frebort, I., Kowalska, M., Hluska, T., Frebortova, J. and Galuszka, P. (2011). Evolution of cytokinin biosynthesis and degradation. *Journal of Experimental Botany* **62**, 2431–2452.
- Freeling, M., Scanlon, M. J. and Fowler, J. E. (2015). ScienceDirectFractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* **35**, 110–118.
- Fricke, W., Akhiyarova, G., Wei, W., Alexandersson, E., Miller, A., Kjellbom, P. O., Richardson, A., T, W., Schreiber, L., Veselov, D. et al. (2006). The short-term growth response to salt of the developing barley leaf. *Journal of Experimental Botany* **57**, 1079–1095.
- Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S. Y., Cutler, S. R., Sheen, J., Rodrigues, P. L. and Zhu, J. K. (2009). In vitro reconstitution of an abscisic acid signalling pathway. *Nature* **462**, 660–664.
- Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L. S. P., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2004). A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal* **39**, 863–876.
- Gagne, J. M., Smalle, J., Gingerich, D. J., Walker, J. M., Yoo, S. D., Yanagisawa, S. and Vierstra, R. D. (2004). Arabidopsis EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress ethylene action and promote growth by directing EIN3 degradation. *Proceedings of the National Academy of Sciences* **101**, 6803–6808.
- Gajdanowicz, P., Michard, E., Sandmann, M., Rocha, M., Correa, L. G. G., Ramirez-Aguilar, S. J., Gomez-Porras, J. L., Gonzalez, W., Thibaud, J. B., van Dongen, J. T. et al. (2011). Potassium (K⁺) gradients serve as a mobile energy source in plant vascular tissues. *Proceedings of the National Academy of Sciences* **108**, 864–869.
- Gallie, D. R. (2015). Ethylene receptors in plants - why so much complexity? *F1000Prime Reports* **7**, 39.

- Galvan-Ampudia, C. S. and Testerink, C.** (2011). Salt stress signals shape the plant root. *Current Opinion in Plant Biology* **14**, 296–302.
- Gan, S. and Amasino, R. M.** (1995). Inhibition of Leaf Senescence by Autoregulated Production of Cytokinin. *Science* **270**, 1986–1988.
- Gao, R., Duan, K., Guo, G., Du, Z., Chen, Z., Li, L., He, T., Lu, R. and Huang, J.** (2013). Comparative Transcriptional Profiling of Two Contrasting Barley Genotypes under Salinity Stress during the Seedling Stage. *International Journal of Genomics* **2013**, 972852.
- Garg, R., Shankar, R., Thakkar, B., Kudapa, H., Krishnamurthy, L., Mantri, N., Varshney, R. K., Bhatia, S. and Jain, M.** (2016). Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. *Nature* **6**, 19228.
- Geilfus, C. M., Zörb, C. and Muhling, K. H.** (2010). Salt stress differentially affects growth-mediating beta-expansins in resistant and sensitive maize (*Zea mays* L.). *Plant Physiology et Biochemistry* **48**, 993–998.
- Geng, Y., Wu, R., Wee, C. W., Xie, F., Wei, X., Chan, P. M. Y., Tham, C., Duan, L. and Dinneny, J. R.** (2013). A Spatio-Temporal Understanding of Growth Regulation during the Salt Stress Response in Arabidopsis. *The Plant Cell* **25**, 2132–2154.
- Genome Project, A.** (2013). Genomic Clues to the Ancestral Flowering Plant. *Science* **342**, 1241089–1124110.
- Ghanem, M. E., Albacete, A., Martinez-Andujar, C., Acosta, M., Romero-Aranda, R., Dodd, I. C., Lutts, S. and Perez-Alfocea, F.** (2008). Hormonal changes during salinity-induced leaf senescence in tomato (*Solanum lycopersicum* L.). *Journal of Experimental Botany* **59**, 3039–3050.
- Ghassemian, M., Nambara, E., Cutler, S., Kawaide, H., Kamiya, Y. and McCourt, P.** (2000). Regulation of Abscissic Acid Signaling by the Ethylene Response Pathway in Arabidopsis. *The Plant Cell* **12**, 1117–1126.
- Ghorbani Javid, M., Sorooshzadeh, A., Modarres Sanavy, S. A. M., Allahdadi, I. and Moradi, F.** (2011). Effects of the exogenous application of auxin and cytokinin on carbohydrate accumulation in grains of rice under salt stress. *Plant Growth Regulation* **65**, 305–313.
- Gilmour, S. J., Fowler, S. G. and Thomashow, M. F.** (2004). Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. *Plant Molecular Biology* **54**, 767–781.
- Gilroy, S., Suzuki, N., Miller, G., Choi, W. G., Toyota, M., Devireddy, A. R. and Mittler, R.** (2014). A tidal wave of signals: calcium and ROS at the forefront of rapid systemicsignaling. *Trends in Plant Science* **19**, 623–630.
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. and Toulmin, C.** (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science* **327**, 812–818.
- Goff, S. A.** (2011). The iPlant collaborative: cyberinfrastructure for plant biology. *Frontiers in Plant Science* **2**, 34.

- Gong, H. J., Randall, D. P. and Flowers, T. J. (2006). Silicon deposition in the root reduces sodium uptake in rice (*Oryza sativa* L.) seedlings by reducing bypass flow. *Plant, Cell and Environment* **29**, 1970–1979.
- Gong, X., Chao, L., Zhou, M., Hong, M., Luo, L., Wang, L., Ying, W., Jingwei, C., Songjie, G. and Fashui, H. (2010). Oxidative damages of maize seedlings caused by exposure to a combination of potassium deficiency and salt stress. *Plant and Soil* **340**, 443–452.
- Gonzalez-Guzman, M., Apostolova, N., Belles, J. M., Barrero, J. M., Piqueras, P., Ponce, M. R., Micol, J. L., Serrano, R. and Rodrigues, P. L. (2002). The Short-Chain Alcohol Dehydrogenase ABA2 Catalyzes the Conversion of Xanthoxin to Absciscic Aldehyde. *The Plant Cell* **14**, 1833–1846.
- Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., Fernández, M. A., Holdsworth, M. J., Perez-Amador, M. A., Kollist, H. et al. (2012). Arabidopsis PYR/PYL/RCAR Receptors Play a Major Role in Quantitative Regulation of Stomatal Aperture and Transcriptional Response to Absciscic Acid. *The Plant Cell* **24**, 2483–2496.
- Gonzalez-Guzman, M., Rodriguez, L., Lorenzo-Orts, L., Pons, C., Sarrion-Perdigones, A., Fernandez, M. A., Peirats-Llobet, M., Forment, J., Moreno-Alvero, M., Cutler, S. R. et al. (2014). Tomato PYR/PYL/RCAR absciscic acid receptors show high expression in root, differential sensitivity to the absciscic acid agonist quinabactin, and the capability to enhance plant drought resistance. *Journal of Experimental Botany* **65**, 4451–4464.
- Gout, J.-F. and Lynch, M. (2015). Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization. *Molecular Biology and Evolution* **32**, 2141–2148.
- Grabov, A. (2007). Plant KT/KUP/HAK Potassium Transporters: Single Family - Multiple Functions. *Annals of Botany* **99**, 1035–1041.
- Grundy, J., Stoker, C. and Carré, I. A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Frontiers in Plant Science* **6**, 648.
- Guilfoyle, T. (2007). Sticking with auxin. *Nature* **446**, 621–622.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2009). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 pp. 1–37.
- Gunes, A., Inal, A., Alpaslan, M., Eraslan, F., Bagci, E. G. and Cicek, N. (2007). Salicylic acid induced changes on some physiological parameters symptomatic for oxidative stress and mineral nutrition in maize (*Zea mays* L.) grown under salinity. *Journal of Plant Physiology* **164**, 728–736.
- Guo, H. and Ecker, J. R. (2003). Plant Responses to Ethylene Gas Are Mediated by SCF^{EBF1/EBF2}-Dependent Proteolysis of EIN3 Transcription Factor. *Cell* **115**, 667–677.
- Guo, P., Baum, M., Grando, S., Ceccarelli, S., Bai, G., Li, R., von Korff, M., Varshney, R. K., Graner, A. and Valkoun, J. (2009). Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *Journal of Experimental Botany* **60**, 3531–3544.

- Gupta, B. and Huang, B.** (2014). Review Article Mechanism of Salinity Tolerance in Plants: Physiological, Biochemical, and Molecular Characterization. *International Journal of Genomics* **2014**, 701596.
- Ha, M., Kim, E. D. and Chen, Z. J.** (2009). Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proceedings of the National Academy of Sciences* **106**, 2295–2300.
- Ha, S., Vankova, R., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.-S. P.** (2012). Cytokinins: metabolism and function in plant adaptation to environmental stresses. *Trends in Plant Science* **17**, 172–179.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D. et al.** (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512.
- Hartmann, L., Pedrotti, L., Weiste, C., Fekete, A., Schierstaedt, J., Göttler, J., Kempa, S., Krischke, M., Dietrich, K., Mueller, M. J. et al.** (2015). Crosstalk between Two bZIP Signaling Pathways Orchestrates Salt-Induced Metabolic Reprogramming in Arabidopsis Roots. *The Plant Cell* **27**, 2244–2260.
- Henry, R. J.** (2014). Genomics strategies for germplasm characterization and the development of climate resilient crops. *Frontiers in Plant Science* **5**, 68.
- Hernandez, J. A., Olmos, E., Corpas, F. J., Sevilla, F. and del Rio, L. A.** (1995). Salt-induced oxidative stress in chloroplasts of pea plants. *Plant Science* **105**, 151–167.
- Hernandez, M., Fernandez-Garcia, N., Diaz-Vivancos, P. and Olmos, E.** (2010). A different role for hydrogen peroxide and the antioxidative system under short and long salt stress in Brassica oleracea roots. *Journal of Experimental Botany* **61**, 521–535.
- Hickman, R., Hill, C., Penfold, C. A., Breeze, E., Bowden, L., Moore, J. D., Zhang, P., Jackson, A., Cooke, E., Bewicke-Copley, F. et al.** (2013). A local regulatory network around three NAC transcription factors in stress responses and senescence in Arabidopsis leaves. *The Plant Journal* **75**, 26–39.
- Hortensteiner, S. and Feller, U.** (2002). Nitrogen metabolism and remobilization during senescence. *Journal of Experimental Botany* **53**, 927–937.
- Hovav, R., Udall, J., Chaudhary, B., Rapp, R., Flagel, R. and Wendel, J. F.** (2008). Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *PNAS* **105**, 6191–6195.
- Howden, S. M., Soussana, J. F., Tubiello, F. N., Chhetri, N., Dunlop, M. and Meinke, H.** (2007). Adapting agriculture to climate change. *Proceedings of the National Academy of Sciences* **104**, 19691–19696.
- Hu, L., Li, H., Chen, L., Lou, Y., Amombo, E. and Fu, J.** (2015). RNA-seq for gene identification and transcript profiling in relation to root growth of bermudagrass (*Cynodon dactylon*) under salinity stress. *BMC Genomics* **16**, 575.

- Hu, Y., Rosa, G. J. M. and Gianola, D.** (2016). Incorporating parent-of-origin effects in whole-genome prediction of complex traits. *Genetics Selection Evolution* pp. 1–15.
- Hu, Y.-z., Zeng, Y.-l., Guan, B. and Zhang, F.-c.** (2011). Overexpression of a vacuolar H⁺-pyrophosphatase and a B subunit of H⁺-ATPase cloned from the halophyte *Halostachys caspica* improves salt tolerance in *Arabidopsis thaliana*. *Plant Cell, Tissue and Organ Culture (PCTOC)* **108**, 63–71.
- Huang, D., Wu, W., Abrams, S. R. and Cutler, A. J.** (2008). The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *Journal of Experimental Botany* **59**, 2991–3007.
- Huang, Z., Tang, J., Duan, W., Wang, Z., Song, X. and Hou, X.** (2015). Molecular evolution, characterization, and expression analysis of SnRK2 gene family in Pak-choi (*Brassica rapa* ssp. *chinensis*). *Frontiers in Plant Science* **6**, 879.
- Hughes, M. E., Hogenesch, J. B. and Kornacker, K.** (2010). JTK CYCLE: An Efficient Non-parametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms* **25**, 372–380.
- Illumina** (2011). HiSeq Sequencing Systems. Technical report.
- Illumina** (2014). RNASample Preparation v2 Guide. Illumina.
- Indorf, M., Cordero, J., Neuhaus, G. and Rodríguez-Franco, M.** (2007). Salt tolerance (STO), a stress-related protein, has a major role in light signalling. *The Plant Journal* **51**, 563–574.
- Ingle, R. A., Stoker, C., Stone, W., Adams, N., Smith, R., Grant, M., Carré, I., Roden, L. C. and Denby, K. J.** (2015). Jasmonate signalling drives time-of-day differences in susceptibility of *Arabidopsis* to the fungal pathogen *Botrytis cinerea*. *The Plant Journal* **84**, 937–948.
- Initiative, T. A. G.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Ishida, K., Yamashino, T. and Mizuno, T.** (2008). Expression of the Cytokinin-Induced Type-A Response Regulator Gene ARR9 Is Regulated by the Circadian Clock in *Arabidopsis thaliana*. *Bioscience, Biotechnology and Biochemistry* **72**, 3025–3029.
- Ismail, A., Riemann, M. and Nick, P.** (2012). The jasmonate pathway mediates salt tolerance in grapevines. *Journal of Experimental Botany* **63**, 2127–2139.
- Issa, R., Barker, G., Marshall, A., Slade, S. E. and Taylor, P.** (2013). An Optimized Method for Profiling Glucosinolate Content in Brassica Enabling Plant Line Selection and Quantitative Trait Locus Mapping. *International Journal of Nutrition and Food Sciences* **2**, 10–16.
- Jaillais, Y. and Chory, J.** (2010). Unraveling the paradoxes of plant hormone signaling integration. *Nature Structural & Molecular Biology* **17**, 642–645.
- Jamil, M., ur Rehman, S., Lee, K. J., Kim, J. M. and Kim, H. S.** (2007). Salinity Reduced Growth PS2 Photochemistry and Chlorophyll Content in Radish. *Sci. Agric. (Piracicaba, Braz.)* **64**, 111–118.

- Jardim-Messeder, D., Caverzan, A., Rauber, R., de Souza Ferreira, E., Margis-Pinheiro, M. and Galina, A.** (2015). Succinate dehydrogenase (mitochondrial complex II) is a source of reactive oxygen species in plants and regulates development and stress responses. *New Phytologist* **208**, 776–789.
- Jayakannan, M., Bose, J., Babourina, O., Rengel, Z. and Shabala, S.** (2013). Salicylic acid improves salinity tolerance in Arabidopsis by restoring membrane potential and preventing salt-induced K⁺ loss via a GORK channel. *Journal of Experimental Botany* **64**, 2255–2268.
- Jiang, Q., Hu, Z., Zhang, H. and Ma, Y.** (2014a). Overexpression of *GmDREB1* improves salt tolerance in transgenic wheat and leaf protein response to high salinity. *The Crop Journal* **2**, 120–131.
- Jiang, S. Y., Ma, A., Ramamoorthy, R. and Ramachandran, S.** (2013a). Genome-Wide Survey on Genomic Variation, Expression Divergence, and Evolution in Two Contrasting Rice Genotypes under High Salinity Stress. *Genome Biology and Evolution* **5**, 2032–2050.
- Jiang, W. K., Liu, Y. L., Xia, E. H. and Gao, L. Z.** (2013b). Prevalent Role of Gene Features in Determining Evolutionary Fates of Whole-Genome Duplication Duplicated Genes in Flowering Plants. *Plant Physiology* **161**, 1844–1861.
- Jiang, X., Leidi, E. O. and Pardo, J. M.** (2010). How do vacuolar NHX exchangers function in plant salt tolerance? *Plant Signaling & Behavior* **5**, 792–795.
- Jiang, X., Leidi, E. O. and Pardo, J. M.** (2014b). How do vacuolar NHX exchangers function in plant salt tolerance? *Plant Signaling & Behavior* **5**, 792–795.
- Jiang, Y. and Deyholos, M. K.** (2006). Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes. *BMC Plant Biology* **6**, 25.
- Jiang, Y. and Deyholos, M. K.** (2008). Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Molecular Biology* **69**, 91–105.
- Jin, J., Zhang, H., Kong, L., Gao, G. and Luo, J.** (2013). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research* **42**, D1182–D1187.
- Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., Garrett, W. M., Kessenbrock, M., Groth, G., Tucker, M. L. et al.** (2012). CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis. *Proceedings of the National Academy of Sciences* **109**, 19486–19491.
- Jung, J.-H., Lee, S., Yun, J., Lee, M. and Park, C.-M.** (2014). Plant Science. *Plant Science* **215–216**, 29–38.
- Kader, A. and Lindberg, S.** (2010). Cytosolic calcium and pH signaling in plants under salinity stress. *Plant Signaling & Behavior* **5**, 233–238.
- Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W. E., Edger, P. P., Links, M. G. et al.** (2014). Polyploid Evolution of the Brassicaceae during the Cenozoic Era. *The Plant Cell* **26**, 2777–2791.
- Kang, J., Park, J., Choi, H., Burla, B., Kretschmar, T., Lee, Y. and Martinoia, E.** (2011). Plant ABC Transporters. *The Arabidopsis Book* **9**, e0153.

- Kang, N. Y., Cho, C., Kim, N. Y. and Kim, J.** (2012). Journal of Plant Physiology. *Journal of Plant Physiology* **169**, 1382–1391.
- Kant, S., Burch, D., Badenhorst, P., Palanisamy, R., Mason, J. and Spangenberg, G.** (2015). Regulated Expression of a Cytokinin Biosynthesis Gene IPT Delays Leaf Senescence and Improves Yield under Rainfed and Irrigated Conditions in Canola (*Brassica napus* L.). *PLoS ONE* **10**, e0116349.
- Kariola, T., Brader, G., Helenius, E., Li, J., Heino, P. and Palva, E. T.** (2006). EARLY RESPONSIVE TO DEHYDRATION 15, a Negative Regulator of Abscissic Acid Responses in Arabidopsis. *Plant Physiology* **142**, 1559–1573.
- Kass, R. E. and Raftery, A. E.** (1995). Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795.
- Kazan, K. and Manners, J. M.** (2012). JAZ repressors and the orchestration of phytohormone crosstalk. *Trends in Plant Science* **17**, 22–31.
- Kepinski, S. and Leyser, O.** (2005). The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* **435**, 446–451.
- Kerr, K. M. and Churchill, G. A.** (2001). Experimental design for gene expression microarrays. *Biostatistics (Oxford, England)* **2**, 183–201.
- Khan, M. I. R., Fatma, M., Per, T. S., Anjum, N. A. and Khan, N. A.** (2015). Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants. *Frontiers in Plant Science* **6**, 462.
- Khan, M. M., Al-Mas'oudi, R., Al-Said, F. and Khan, I.** (2013). Salinity Effects on Growth, Electrolyte Leakage, Chlorophyll Content and Lipid Peroxidation in Cucumber (*Cucumis sativus* L.). *2013 International Conference on Food and Agricultural Sciences IPCBEE* **5**, 28–32.
- Khan, N., Syeed, S., Masood, A., Nazar, R. and Iqbal, N.** (2010). Application of salicylic acid increases contents of nutrients and antioxidative metabolism in mungbean and alleviates adverse effects of salinity stress. *International Journal of Plant Biology* **1**, e1.
- Kiegle, E., Moore, C. A., Haseloss, J., Tester, M. A. and Knight, M. R.** (2000). Cell-type-specific calcium responses to drought, salt and cold in the Arabidopsis root. *The Plant Journal* **23**, 267–278.
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J. and Harter, K.** (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *The Plant Journal* **50**, 347–363.
- Kim, H. J., Hong, S. H., Kim, Y. W., Lee, I. H., Jun, J. H., Phee, B. K., Rupak, T., Jeong, H., Lee, Y., Hong, B. S. et al.** (2014). Gene regulatory cascade of senescence-associated NAC transcription factors activated by ETHYLENE-INSENSITIVE2-mediated leaf senescence signalling in Arabidopsis. *Journal of Experimental Botany* **65**, 4023–4036.

- Kim, J. I., Baek, D., Park, H. C., Chun, H. J., Oh, D.-H., Lee, M. K., Cha, J.-Y., Kim, W.-Y., Kim, M. C., Chung, W. S. et al.** (2013). Overexpression of Arabidopsis YUCCA6 in Potato Results in High-Auxin Developmental Phenotypes and Enhanced Resistance to Water Deficit. *Molecular Plant* **6**, 337–349.
- Kim, J. M., Sasaki, T., Ueda, M., Sako, K. and Seki, M.** (2015). Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Frontiers in Plant Science* **6**, 114.
- King, O. D., Foulger, R. E., Dwight, S. S., White, J. V. and Roth, F. P.** (2003). Predicting Gene Function From Patterns of Annotation. *Genome Research* **13**, 896–904.
- Kirk, P., Griffin, J. E., Savage, R. S., Ghahramani, Z. and Wild, D. L.** (2012). Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics* **28**, 3290–3297.
- Kiyosue, T., Yamaguchi-Shinozaki, K. and Shinozaki, K.** (1994). ERD15, a cDNA for a Dehydration-Induced Gene from Arabidopsis thaliana. *Plant Physiology* **106**, 1707.
- Knight, H., Trewavas, A. J. and Knight, M. R.** (1997). Calcium signalling in Arabidopsis thaliana responding to drought and salinity. *The Plant Journal* **12**, 1067–1078.
- Korlach, J.** (2013). Understanding Accuracy in SMRT. *Pacific Biosciences* pp. 1–9.
- Kosma, D. K., Bourdenx, B., Bernard, A., Parsons, E. P., Lu, S., Joubes, J. and Jenks, M. A.** (2009). The Impact of Water Deficiency on Leaf Cuticle Lipids of Arabidopsis. *Plant Physiology* **151**, 1918–1929.
- Kovacs, D., Kalmar, E., Torok, Z. and Tompa, P.** (2008). Chaperone Activity of ERD10 and ERD14, Two Disordered Stress-Related Plant Proteins. *Plant Physiology* **147**, 381–390.
- Krishnamurthy, P., Jyothi-Prakash, P. A., Qin, L., He, J., Lin, Q., Loh, C. S. and Kumar, K. K.** (2014). Role of root hydrophobic barriers in salt exclusion of a mangrove plant Avicennia officinalis. *Plant, Cell and Environment* **37**, 1656–1671.
- Krishnamurthy, P., Ranathunge, K., Nayak, S., Schreiber, L. and Mathew, M. K.** (2011). Root apoplastic barriers block Na⁺ transport to shoots in rice (Oryza sativa L.). *Journal of Experimental Botany* **62**, 4215–4228.
- Kugler, A., Köhler, B., Palme, K., Wolff, P. and Dietrich, P.** (2009). Salt-dependent regulation of a CNG channel subfamily in Arabidopsis. *BMC Plant Biology* **9**, 140.
- Kumar, G., Purty, R. S., Sharma, M. P., Singla-Pareek, S. L. and Pareek, A.** (2009). Physiological responses among Brassica species under salinity stress show strong correlation with transcript abundance for SOS pathway-related genes. *Journal of Plant Physiology* **166**, 507–520.
- Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., Fukuda, H., Sugimoto, K. and Sakakibara, H.** (2009). Functional Analyses of LONELY GUY Cytokinin-Activating Enzymes Reveal the Importance of the Direct Activation Pathway in Arabidopsis. *The Plant Cell* **21**, 3152–3169.
- Kurusu, T., Kuchitsu, K. and Tada, Y.** (2015). Plant signaling networks involving Ca²⁺ and Rboh/Nox-mediated ROS production under salinity stress. *Frontiers in Plant Science* **6**, 427.

- Labaj, P. P., Lepar, G. G., Linggi, B. E., Markillie, L. M., Wiley, H. S. and Kreil, D. P.** (2011). Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics* **27**, i383–i391.
- Laird, J., Armengaud, P., Giuntini, P., Laval, V. r. and Milner, J. J.** (2004). Inappropriate annotation of a key defence marker in Arabidopsis: will the real PR-1 please stand up? *Planta* **219**, 1089–1092.
- Lata, C. and Prasad, M.** (2011). Role of DREBs in regulation of abiotic stress responses in plants. *Journal of Experimental Botany* **62**, 4731–4748.
- Lawson, H. A., Cheverud, J. M. and Wolf, J. B.** (2013). Genomic imprinting and parent- of-origin effects on complex traits. *Nature Reviews Genetics* **14**, 609–617.
- Le, D. T., Nishiyama, R., Watanabe, Y., Vankova, R., Tanaka, M., Seki, M., Ham, L. H., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.-S. P.** (2012). Identification and Expression Analysis of Cytokinin Metabolic Genes in Soybean under Normal and Drought Conditions in Relation to Cytokinin Levels. *PLoS ONE* **7**, e42411.
- Leach, L. J., Belfield, E. J., Jiang, C., Brown, C., Mithani, A. and Harberd, N. P.** (2014). Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics* **15**, 276.
- Lee, E. K., Kwon, M., Ko, J. H., Yi, H., Hwang, M. G., Chang, S. and Cho, M. H.** (2004). Binding of Sulfonylurea by AtMRP5, an Arabidopsis Multidrug Resistance-Related Protein That Functions in Salt Tolerance. *Plant Physiology* **134**, 528–538.
- Leitch, A. R. and Leitch, I. J.** (2008). Genomic Plasticity and the Diversity of Polyploid Plants. *Science* **320**, 481–483.
- Leivar, P., Monte, E., Al-Sady, B., Carle, C., Storer, A., Alonso, J. M., Ecker, J. R. and Quail, P. H.** (2008). The Arabidopsis Phytochrome-Interacting Factor PIF7, Together with PIF3 and PIF4, Regulates Responses to Prolonged Red Light by Modulating phyB Levels. *The Plant Cell* **20**, 337–352.
- LeProust, E.** (2015). Agilent’s Microarray Platform: How High-Fidelity DNA Synthesis Maximizes the Dynamic Range of Gene Expression Measurements. Technical report.
- Leung, J., Merlot, S. and Giraudat, J.** (2002). The Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 Genes Encode Homologous Protein Phosphatases 2C Involved in Absciscic Acid Signal Transduction. *The Plant Cell* **9**, 759–771.
- Lewis, L. A., Polanski, K., de Torres-Zabala, M., Jayaraman, S., Bowden, L., Moore, J., Penfold, C. A., Jenkins, D. J., Hill, C., Baxter, L. et al.** (2015). Transcriptional Dynamics Driving MAMP-Triggered Immunity and Pathogen Effector-Mediated Immunosuppression in Arabidopsis Leaves Following Infection with *Pseudomonas syringae* pv tomato DC3000. *The Plant Cell* **27**, 3038–3064.
- Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., Wu, J. et al.** (2014a). mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *The Plant Cell* **26**, 1878–1900.

- Li, C., Ng, C. K. Y. and Fan, L.-M.** (2015a). Environmental and Experimental Botany. *Environmental and Experimental Botany* **114**, 80–91.
- Li, C., Potuschak, T., Colon-Carmona, A., Gutierrez, R. and Doerner, P.** (2005). Arabidopsis TCP20 links regulation of growth and cell division control pathways. *Proceedings of the National Academy of Sciences* **102**, 36.
- Li, G., Peng, X., Wei, L. and Kang, G.** (2013). Short Communication. *Gene* **529**, 321–325.
- Li, H., Sun, J., Xu, Y., Jiang, H., Wu, X. and Li, C.** (2007). The bHLH-type transcription factor AtAIB positively regulates ABA response in Arabidopsis. *Plant Molecular Biology* **65**, 655–665.
- Li, J., Gao, G., Xu, K., Chen, B., Yan, G., Li, F., Qiao, J., Zhang, T. and Wu, X.** (2014b). Genome-Wide Survey and Expression Analysis of the Putative Non-Specific Lipid Transfer Proteins in Brassica rapa L. *PLoS ONE* **9**, e84556.
- Li, J., Liu, B., Cheng, F., Wang, X., Aarts, M. G. M. and Wu, J.** (2014c). Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in Brassica rapa. *New Phytologist* **203**, 182–194.
- Li, S., Fu, Q., Chen, L., Huang, W. and Yu, D.** (2011). Arabidopsis thaliana WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta* **233**, 1237–1252.
- Li, S., Zhou, X., Chen, L., Huang, W. and Yu, D.** (2010). Functional characterization of Arabidopsis thaliana WRKY39 in heat stress. *Molecules and Cells* **29**, 475–483.
- Li, S.-B., Xie, Z.-Z., Hu, C.-G. and Zhang, J.-Z.** (2016). A Review of Auxin Response Factors (ARFs) in Plants. *Frontiers in Plant Science* **7**, 47.
- Li, W. and Godzik, A.** (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.
- Li, W., Zhao, F., Fang, W., Xie, D., Hou, J., Yang, X., Zhao, Y., Tang, Z., Nie, L. and Lv, S.** (2015b). Identification of early salt stress responsive proteins in seedling roots of upland cotton (*Gossypium hirsutum* L.) employing iTRAQ-based proteomic technique. *Frontiers in Plant Science* **6**, 732.
- Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., Guo, J., Liang, W., Chen, L., Yin, J. et al.** (2006). Genome-Wide Analysis of Basic/Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis. *Plant Physiology* **141**, 1167–1184.
- Lim, C., Baek, W., Jung, J., Kim, J.-H. and Lee, S.** (2015). Function of ABA in Stomatal Defense against Biotic and Drought Stresses. *International Journal of Molecular Sciences* **16**, 15251–15270.
- Liu, C., Zhang, X., Zhang, K., An, H., Hu, K., Wen, J., Shen, J., Ma, C., Yi, B., Tu, J. et al.** (2015a). Comparative Analysis of the Brassica napus Root and Leaf Transcript Profiling in Response to Drought Stress. *International Journal of Molecular Sciences* **16**, 18752–18777.
- Liu, F., Jensen, C. R., Shahanzari, A., Andersen, M. N. and Jacobsen, S.-E.** (2005). ABA regulated stomatal control and photosynthetic water use efficiency of potato (*Solanum tuberosum* L.) during progressive soil drying. *Plant Science* **168**, 831–836.

- Liu, J., Ishitani, M., Halfter, U., Kim, C. S. and Zhu, J. K.** (2000). The *Arabidopsis thaliana* SOS2 gene encodes a protein kinase that is required for salt tolerance. *Proceedings of the National Academy of Sciences* **97**, 3730–3734.
- Liu, J. X. and Howell, S. H.** (2010). bZIP28 and NF-Y Transcription Factors Are Activated by ER Stress and Assemble into a Transcriptional Complex to Regulate Stress Response Genes in *Arabidopsis*. *The Plant Cell* **22**, 782–796.
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., Zhao, M., Ma, J., Yu, J., Huang, S. et al.** (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature* **5**, 3930.
- Liu, Y., Du, H., Wang, K., Huang, B. and Wang, Z.** (2011). Differential Photosynthetic Responses to Salinity Stress between Two Perennial Grass Species Contrasting in Salinity Tolerance. *HortScience* **46**, 311–316.
- Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y. and Sun, Q.** (2015b). Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biology* **15**, 152.
- Llorca, C. M., Potschin, M. and Zentgraf, U.** (2014). bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. *Frontiers in Plant Science* **5**, 169.
- Lou, P., Wu, J., Cheng, F., Cressman, L. G., Wang, X. and McClung, C. R.** (2012). Preferential Retention of Circadian Clock Genes during Diploidization following Whole Genome Triplication in *Brassica rapa*. *The Plant Cell* **24**, 2415–2426.
- Love, C. G., Graham, N. S., Ó Lochlainn, S., Bowen, H. C., May, S. T., White, P. J., Broadley, M. R., Hammond, J. P. and King, G. J.** (2010). A *Brassica* Exon Array for Whole-Transcript Gene Expression Profiling. *PLoS ONE* **5**, e12812.
- Love, M. I., Huber, W. and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.
- Lu, X., Guan, Q. and Zhu, J.** (2014). Downregulation of CSD2 by a heat-inducible miR398 is required for thermotolerance in *Arabidopsis*. *Plant Signaling & Behavior* **8**, e24952.
- Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A. and Grill, E.** (2009). Regulators of PP2C Phosphatase Activity Function as Abscissic Acid Sensors. *Science* **324**, 1064–1068.
- Maathuis, F. J. M.** (2013). Sodium in plants: perception, signalling, and regulation of sodium fluxes. *Journal of Experimental Botany* **65**, 849–858.
- Maathuis, F. J. M., Ahmad, I. and Patishtan, J.** (2014). Regulation of Na⁺ fluxes in plants. *Frontiers in Plant Science* **5**, 467.
- Maathuis, F. J. M. and Amtmann, A.** (1999). K⁺ Nutrition and Na⁺ Toxicity: The Basis of Cellular K⁺/Na⁺ Ratios. *Annals of Botany* **84**, 123–133.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y.** (2005a). Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences* **12**, 5454–5459.

- Maere, S., Heymans, K. and Kuiper, M.** (2005b). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**, 3448–3449.
- Maggioni, L.** (2015). *Domestication of Brassica oleracea L.* Ph.D. thesis, Swedish University of Agricultural Sciences.
- Mano, Y. and Nemoto, K.** (2012). The pathway of auxin biosynthesis in plants. *Journal of Experimental Botany* **63**, 2853–2872.
- Marschner, H.** (1995). *Mineral Nutrition of Higher Plants*. Elsevier.
- Martin, L. B. B., Fei, Z., Giovannoni, J. J. and Rose, J. K. C.** (2013). Catalyzing plant science research with RNA-seq. *Frontiers in Plant Science* **4**, 66.
- Martin-Magniette, M.-L., Aubert, J., Cabannes, E. and Daudin, J.-J.** (2005). Evaluation of the gene-specific dye bias in cDNA microarray experiments. *Bioinformatics* **21**, 1995–2000.
- Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2004). Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. *The Plant Journal* **38**, 982–993.
- Maser, P., Thomine, S., Schroeder, J. I., Ward, J. M., Hirschi, K., Sze, H., Talke, I. N., Amtmann, A., Maathuis, F. J. M., Sanders, D. et al.** (2001). Phylogenetic Relationships within Cation Transporter Families of Arabidopsis. *Plant Physiology* **126**, 1646–1667.
- Mason, S. A., Sayyid, F., Kirk, P. D. W., Starr, C. and Wild, D. L.** (2016). MDI-GPU: accelerating integrative modelling for genomic-scale data using GP-GPU computing. *Statistical Applications in Genetics and Molecular Biology* **15**, 83–86.
- McHattie, S.** (2011). *Modelling transcriptional networks in plant senescence*. Ph.D. thesis, University of Warwick.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R.** (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* **41**, W597–W600.
- Memon, S. A., Hou, X. and Wang, L. J.** (2010). Morphological Analysis of Salt Stress Response of Pak Choi. *EJEAFChe* **9**, 248–254.
- Mengiste, T., Chen, X., Salmeron, J. and Dietrich, R.** (2003). The BOTRYTIS SUSCEPTIBLE1 Gene Encodes an R2R3MYB Transcription Factor Protein That Is Required for Biotic and Abiotic Stress Responses in Arabidopsis. *The Plant Cell* **15**, 2551–2565.
- Merwad, A. R. M. A.** (2016). Efficiency of Potassium Fertilization and Salicylic Acid on Yield and Nutrient Accumulation of Sugar Beet Grown on Saline Soil. *Communications in Soil Science and Plant Analysis* **47**, 1184–1192.
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H. and Michelmore, R. W.** (2003). Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis. *The Plant Cell* **15**, 809–834.

- Mhamdi, A., Queval, G., Chaouch, S., Vanderauwera, S., Van Breusegem, F. and Noctor, G. (2010). Catalase function in plants: a focus on Arabidopsis mutants as stress-mimic models. *Journal of Experimental Botany* **61**, 4197–4220.
- Mian, A., Oomen, R. J. F. J., Isayenkov, S., Sentenac, H., Maathuis, F. J. M. and Véry, A.-A. (2011). Over-expression of an Na⁺- and K⁺-permeable HKT transporter in barley improves salt tolerance. *The Plant Journal* **68**, 468–479.
- Miller, G., Suzuki, N., Ciftci-Yilmaz, S. and Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant, Cell and Environment* **33**, 453–467.
- Mitsuya, S., El-Shami, M., Sparkes, I. A., Charlton, W. L., De Marcos Lousa, C., Johnson, B. and Baker, A. (2010). Salt Stress Causes Peroxisome Proliferation, but Inducing Peroxisome Proliferation Does Not Improve NaCl Tolerance in Arabidopsis thaliana. *PLoS ONE* **5**, e9408.
- Mittal, S., Kumari, N. and Sharma, V. (2012). Plant Physiology and Biochemistry. *Plant Physiology et Biochemistry* **54**, 17–26.
- Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., Lee, H., Stevenson, B. and Zhu, J.-K. (2006). Gain- and loss-of-function mutations in *Zat10* enhance the tolerance of plants to abiotic stress. *FEBS Letters* **580**, 6537–6542.
- Mittler, R., Vanderauwera, S., Gollery, M. and Van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends in Plant Science* **9**, 490–498.
- Miura, K. and Yasuomi, T. (2014). Regulation of water, salinity, and cold stress responses by salicylic acid. *Frontiers in Plant Science* **5**, 4.
- Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2012). Biochimica et Biophysica Acta. *BBA - Gene Regulatory Mechanisms* **1819**, 86–96.
- Moffat, C. S., Ingle, R. A., Wathugala, D. L., Saunders, N. J., Knight, H. and Knight, M. R. (2012). ERF5 and ERF6 Play Redundant Roles as Positive Regulators of JA/Et-Mediated Defense against Botrytis cinerea in Arabidopsis. *PLoS ONE* **7**, e35995.
- Moghe, G. D., Hufnagel, D. E., Tang, H., Xiao, Y., Dworkin, I., Town, C. D., Conner, J. K. and Shiu, S. H. (2014). Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish *Raphanus raphanistrum* and Three Other Brassicaceae Species. *The Plant Cell* **26**, 1925–1937.
- Muller, P., Li, X. P. and Niyogi, K. K. (2001). Non-Photochemical Quenching. A Response to Excess Light Energy. *Plant Physiology* **125**, 1558–1566.
- Mun, J.-H., Yu, H.-J., Shin, J. Y., Oh, M., Hwang, H.-J. and Chung, H. (2012). Auxin response factor gene family in Brassica rapa: genomic organization, divergence, expression, and evolution. *Molecular Genetics and Genomics* **287**, 765–784.
- Munns, R. and Tester, M. (2008). Mechanisms of Salinity Tolerance. *Annual Review of Plant Biology* **59**, 651–681.

- Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., Steinbach, D., Quesneville, H., Cooke, R. and Salse, J.** (2014). Shared Subgenome Dominance Following Polyploidization Explains Grass Genome Evolutionary Plasticity from a Seven Protochromosome Ancestor with 16K Protogenes. *Genome Biology and Evolution* **6**, 12–33.
- Nagano, Y., Furuhashi, H., Inaba, T. and Sasaki, Y.** (2001). A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Research* **29**, 4097–4105.
- Nakaminami, K., Karlson, D. T. and R, I.** (2006). Functional conservation of cold shock domains in bacteria and higher plants. *Proceedings of the National Academy of Sciences* **103**, 26.
- Nakashima, K., Shinwari, Z. K., Sakuma, Y., Seki, M., Miura, S., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2000). Organization and expression of two *Arabidopsis DREB2* genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. *Plant Molecular Biology* **42**, 657–665.
- Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2012). NAC transcription factors in plant abiotic stress responses. *BBA - Gene Regulatory Mechanisms* **1819**, 97–103.
- Nakashima, K. and Yamaguchi-Shinozaki, K.** (2013). ABA signaling in stress-response and seed development. *Plant Cell Reports* **32**, 959–970.
- Nakashima, K., Yamaguchi-Shinozaki, K. and Shinozaki, K.** (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Frontiers in Plant Science* **5**, 170.
- Nambara, E. and Marion-Poll, A.** (2005). Absciscic Acid Biosynthesis and Catabolism. *Annual Review of Plant Biology* **56**, 165–185.
- Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. *The Plant Journal* **34**, 137–148.
- Nations, U.** (2014). Concise Report on the World Population Situation in 2014. *Economic and Social Affairs, United Nations* pp. 1–38.
- Ndimba, B. K., Chivasa, S., Simon, W. J. and Slabas, A. R.** (2005). Identification of *Arabidopsis* salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry. *PROTEOMICS* **5**, 4185–4196.
- Nelson, M. and Mareida, M.** (2001). Environmental impacts of the CGIAR: an assessment, in Doc. No. SDR/TAC:IAR/01/11 presented to the Mid-Term Meeting. 21-25 May, Durban, South Africa. In *Doc. No. SDR/TAC:IAR/01/11 presented to the Mid-Term Meeting. 21-25 May, Durban, South Africa.*
- Ng, L. M., Melcher, K., Teh, B. T. and Xu, H. E.** (-2013). Absciscic acid perception and signaling: structural mechanisms and applications. *Nature* **35**, 567–584.

- Nguyen, C. T., Agorio, A., Jossier, M., Depré, S., Thomine, S. and Filleur, S. (2016a). Characterization of the Chloride Channel-Like, AtCLCg, Involved in Chloride Tolerance in *Arabidopsis thaliana*. *Plant and Cell Physiology* **57**, 764–775.
- Nguyen, K. H., Ha, C. V., Nishiyama, R., Watanabe, Y., Leyva-González, M. A., Fujita, Y., Tran, U. T., Li, W., Tanaka, M., Seki, M. et al. (2016b). Arabidopsistype B cytokinin response regulators ARR1, ARR10, and ARR12 negatively regulate plant responses to drought. *Proceedings of the National Academy of Sciences* **113**, 201600399.
- Nieto-Díaz, M., Pita-Thomas, W. and Nieto-Sampedro, M. (2007). Cross-species analysis of gene expression in non-model mammals: reproducibility of hybridization on high density oligonucleotide microarrays. *BMC Genomics* **8**, 89.
- Nishiyama, R., Watanabe, Y., Fujita, Y., Le, D. T., Kojima, M., Werner, T., Vankova, R., Yamaguchi-Shinozaki, K., Shinozaki, K., Kakimoto, T. et al. (2012). Analysis of Cytokinin Mutants and Regulation of Cytokinin Metabolic Genes Reveals Important Regulatory Roles of Cytokinins in Drought, Salt and Abscissic Acid Responses, and Abscissic Acid Biosynthesis. *The Plant Cell* **23**, 2169–2183.
- Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlen, M. and Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **40**, 10084–10097.
- Noreen, S., Ashraf, M., Hussain, M. and Jamil, A. (2009). Exogenous application of salicylic acid enhances antioxidative capacity in salt stressed sunflower (*Helianthus annuus* L.) plants. *Pakistan Journal of Botany* **41**, 473–479.
- O'Brian, J. A. and Benková, E. (2013). Cytokinin cross-talking during biotic and abiotic stress responses. *Frontiers in Plant Science* **4**, 451.
- Oh, J. E., Kwon, Y., Kim, J. H., Noh, H., Hong, S. W. and Lee, H. (2011). A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress. *Plant Molecular Biology* **77**, 91–103.
- Oh, S. J., Song, S. I., Kim, Y. S., Jang, H. J., Kim, S. Y., Kim, M., Kim, Y. K., Nahm, B. H. and Kim, J. K. (2005). *Arabidopsis* CBF3/DREB1A and ABF3 in Transgenic Rice Increased Tolerance to Abiotic Stress without Stunting Growth. *Plant Physiology* **138**, 341–351.
- Ohno, S. (1970). *Evolution by gene duplication*. Berlin: Springer-Verlag.
- Olias, R., Eljakaoui, Z., Li, J., De Morales, P. A., Marin-Manzano, M. C., Pardo, J. M. and Belver, A. (2009). The plasma membrane Na⁺/H⁺ antiporter SOS1 is essential for salt tolerance in tomato and affects the partitioning of Na⁺ between plant organs. *Plant, Cell and Environment* **32**, 904–916.
- Omidbakhshfard, M. A., Omranian, N., Ahmadi, F. S., nikoloski, Z. and Mueller-Roeber, B. (2012). Effect of salt stress on genes encoding translation-associated proteins in *Arabidopsis thaliana*. *Plant Signaling & Behavior* **7**, 1095–1102.

- Onkokesung, N., Galis, I., von Dahl, C. C., Matsuoka, K., Saluz, H. P. and Baldwin, I. T. (2010). Jasmonic Acid and Ethylene Modulate Local Responses to Wounding and Simulated Herbivory in *Nicotiana attenuata* Leaves. *Plant Physiology* **153**, 785–798.
- Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P. et al. (2003). Comprehensive Analysis of NAC Family Genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research* **10**, 239–247.
- Osakabe, Y., Miyata, S., Urao, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2002). Overexpression of *Arabidopsis* response regulators, ARR4/ATRR1/IBC7 and ARR8/ATRR3, alters cytokinin responses differentially in the shoot and in callus formation. *Biochemical and Biophysical Research Communications* **293**, 806–815.
- Padawer, T., Leighty, R. E. and Wang, D. (2012). Duplicate gene enrichment and expression pattern diversification in multicellularity. *Nucleic Acids Research* **40**, 7597–7605.
- Pandey, N., Ranjan, A., Pant, P., Tripathi, R. K., Ateek, F., Pandey, H. P., Patre, U. V. and Sawant, S. V. (2013). CAMTA 1 regulates drought responses in *Arabidopsis thaliana*. *BMC Genomics* **14**, 216.
- Paritosh, K., Yavada, S. K., Gupta, V., Panjab-Massand, P., Sodhi, Y. S., Pradhan, A. K. and Pental, D. (2013). RNA-seq based SNPs in some agronomically important oleiferous lines of *Brassica rapa* and their use for genome-wide linkage mapping and specific-region fine mapping. *BMC Genomics* **14**, 463.
- Park, S. Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T. f. F. et al. (2009). Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science* **324**, 1068–1071.
- Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., Town, C. D., Nixon, J., Krishnakumar, V., Bidwell, S. L. et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology* **15**, R77.
- Pasapula, V., Shen, G., Kuppu, S., Paez-Valencia, J., Mendoza, M., Hou, P., Chen, J., Qiu, X., Zhu, L., Zhang, X. et al. (2010). Expression of an *Arabidopsis* vacuolar H⁺-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. *Plant Biotechnology Journal* **9**, 88–99.
- Paterson, A. H., Chapman, B. A., Kissinger, J. C., Bowers, J. E., Feltus, F. A. and Estill, J. C. (2006). Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends in Genetics* **22**, 597–602.
- Patterson, T. A., Lobenhofer, E. K., Fulmer-Smentek, S. B., Collins, P. J., Chu, T.-M., Bao, W., Fang, H., Kawasaki, E. S., Hager, J., Tikhonova, I. R. et al. (2006). Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project. *Nature Biotechnology* **24**, 1140–1150.

- Pauwels, L., Barbero, G. F., Geerinck, J., Tilleman, S., Grunewald, W., Pérez, A. C., Chico, J. M., Bossche, R. V., Sewell, J., Gil, E. et al. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* **464**, 788–791.
- Pedranzani, H., Racagni, G., Alemano, S., Miersch, O., Ramirez, I., Pena-Cortes, H., Taleisnik, E., Machado-Domenech, E. and Abdala, G. (2003). Salt tolerant tomato plants show increased levels of jasmonic acid. *Plant Growth Regulation* **41**, 149–158.
- Peleg, Z. and Blumwald, E. (2011). Hormone balance and abiotic stress tolerance in crop plants. *Current Opinion in Plant Biology* **14**, 290–295.
- Peleg, Z., Reguera, M., Tumimbang, E., Walia, H. and Blumwald, E. (2011). Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. *Plant Biotechnology Journal* **9**, 747–758.
- Penfold, C. A. and Buchanan-Wollaston, V. (2014). Modelling transcriptional networks in leaf senescence. *Journal of Experimental Botany* **65**, 3859–3873.
- Penfold, C. A., Buchanan-Wollaston, V., Denby, K. J. and Wild, D. L. (2012). Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics* **28**, i233–i241.
- Penfold, C. A., Shifaz, A., Brown, P. E., Nicholson, A. and Wild, D. L. (2015). CSI: a nonparametric Bayesian approach to network inference from multiple perturbed time series gene expression data. *Statistical Applications in Genetics and Molecular Biology* **14**, 307–310.
- Penfold, C. A. and Wild, D. L. (2011). How to infer gene networks from expression profiles, revisited. *Interface Focus* **1**, 857–870.
- Peng, Z., He, S., Gong, W., Sun, J., Pan, Z., Xu, F., Lu, Y. and Du, X. (2014). Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics* **15**, 760.
- Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G. G., Rensing, S. A., Kersten, B. and Mueller-Roeber, B. (2009). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Research* **38**, D822–D827.
- Petersson, S. V., Johansson, A. I., Kowalczyk, M., Makoveychuk, A., Wang, J. Y., Moritz, T., Grebe, M., Benfey, P. N., Sandberg, G. and Ljung, K. (2009). An Auxin Gradient and Maximum in the Arabidopsis Root Apex Shown by High-Resolution Cell-Specific Analysis of IAA Distribution and Synthesis. *The Plant Cell* **21**, 1659–1668.
- Petricka, J. J., Winter, C. M. and Benfey, P. N. (2012). Control of Arabidopsis Root Development. *Annual Review of Plant Biology* **63**, 563–590.
- Picarella, M. E., Antonelli, M. G., Astolfi, S., Zuchi, S., Vernieri, P. and Soressi, G. P. (2007). Ethylene and ABA cross-communication and plant growth response to salt stress in tomato (*Solanum lycopersicum* L.). In *Advances in Plant Ethylene Research: Proceedings of the 7th International Symposium on the Plant Hormone Ethylene*, pp. 1–2. Springer.

- Pink, D., Bailey, L., McClement, S., Hand, P., Mathas, E., Buchanan-Wollaston, V., Astley, D., King, G. and Teakle, G. (2008). Double haploids, markers and QTL analysis in vegetable brassicas. *Euphytica* **164**, 509–514.
- Pinoli, P., Chicco, D. and Masseroli, M. (2015). Computational algorithms to predict Gene Ontology annotations. *BMC Bioinformatics* **16**, S4.
- Pokhilko, A., ndez, A. P. n. a. F. a., Edwards, K. D., Southern, M. M., Halliday, K. J. and Millar, A. J. (2012). The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Molecular Systems Biology* **8**, 574.
- Polanski, K., Rhodes, J., Hill, C., Zhang, P., Jenkins, D. J., Kiddle, S., Jironkin, A., Beynon, J., Buchanan-Wollaston, V., Ott, S. et al. (2014). Wigwags: identifying gene modules co-regulated across multiple biological conditions. *Bioinformatics* **30**, 962–970.
- Pont, C., Murat, F., Guizard, S., Flores, R., Foucrier, S., Bidet, Y., Quraishi, U. M., Alaux, M., Doležel, J., Fahima, T. et al. (2013). Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *The Plant Journal* **76**, 1030–1044.
- Potuschak, T., Lechner, E., Parmentier, Y., Yanagisawa, S., Grava, S., Koncz, C. and Genschik, P. (2003). EIN3-Dependent Regulation of Plant Ethylene Hormone Signaling by Two Arabidopsis F Box Proteins: EBF1 and EBF2. *Cell* **115**, 679–689.
- Puranik, S., Sahu, P. P., Srivastava, P. S. and Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. *Trends in Plant Science* **17**, 369–381.
- Qadir, M., Quillerou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R. J., Drechsel, P. and Noble, A. D. (2014). Economics of salt-induced land degradation and restoration. *Natural Resources Forum* **38**, 282–295.
- Qados, A. M. S. A. (2011). Effect of salt stress on plant growth and metabolism of bean plant *Vicia faba* (L.). *Journal of the Saudi Society of Agricultural Sciences* **10**, 7–15.
- Qian, W., Liao, B.-Y., Chang, C. and Zhang, J. (2010). Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* **26**, 425–430.
- Qiao, H., Chang, K. N., Yazaki, J. and Ecker, J. R. (2009). Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. *Genes & Development* **23**, 512–521.
- Qiao, H., Shen, Z., Huang, S. s. C., Schmitz, R. J., Urich, M. A., Briggs, S. P. and Ecker, J. R. (2012). Processing and Subcellular Trafficking of ER-Tethered EIN2 Control Response to Ethylene Gas. *Science* **338**, 390–393.
- Qu, C., Liu, C., Gong, X., Li, C., Hong, M., Wang, L. and Hong, F. (2012). Environmental and Experimental Botany. *Environmental and Experimental Botany* **75**, 134–141.
- Qu, C., Liu, C., Ze, Y., Gong, X., Hong, M., Wang, L. and Hong, F. (2011). Inhibition of Nitrogen and Photosynthetic Carbon Assimilation of Maize Seedlings by Exposure to a Combination of Salt Stress and Potassium-Deficient Stress. *Biological Trace Element Research* **144**, 1159–1174.

- Quesada, V., García-Martínez, S., Piqueras, P., Ponce, M. R. and Micol, J. L.** (2002). Genetic Architecture of NaCl Tolerance in Arabidopsis. *Plant Physiology* **130**, 951–963.
- Raghavendra, A. S., Gonugunta, V. K., Christmann, A. and Grill, E.** (2010). ABA perception and signalling. *Trends in Plant Science* **15**, 395–401.
- Ranjit, S. L., Manish, P. and Penna, S.** (2016). Early osmotic, antioxidant, ionic, and redox responses to salinity in leaves and roots of Indian mustard (*Brassica juncea* L.). *Protoplasma* **253**, 101–110.
- Rashotte, A. M. and Goertzen, L. R.** (2015). The CRF domain defines Cytokinin Response Factor proteins in plants. *BMC Plant Biology* **10**, 74.
- Ré, D. A., Capella, M., Bonaventure, G. and Chan, R. L.** (2014). Arabidopsis AtHB7 and AtHB12 evolved divergently to fine tune processes associated with growth and responses to water stress. *BMC Plant Biology* **14**, 150.
- Reeves, P. A., Panella, L. W. and Richards, C. M.** (2012). Retention of agronomically important variation in germplasm core collections: implications for allele mining. *Theoretical and Applied Genetics* **124**, 1155–1171.
- Remy, E. and Duque, P.** (2014). Beyond cellular detoxification: a plethora of physiological roles for MDR transporter homologs in plants. *Frontiers in Physiology* **5**, 201.
- Ren, G., An, K., Liao, Y., Zhou, X., Cao, Y., Zhao, H., Ge, X. and Kuai, B.** (2007). Identification of a Novel Chloroplast Protein AtNYE1 Regulating Chlorophyll Degradation during Leaf Senescence in Arabidopsis. *Plant Physiology* **144**, 1429–1441.
- Rengasamy, P.** (2006). World salinization with emphasis on Australia. *Journal of Experimental Botany* **57**, 1017–1023.
- Riemann, M., Dhakarey, R., Hazman, M., Miro, B., Kohli, A. and Nick, P.** (2015). Exploring Jasmonates in the Hormonal Network of Drought and Salinity Responses. *Frontiers in Plant Science* **6**, 1077.
- Rivero, R. M., Kojima, M., Gepstein, A., Sakakibara, H., Mittler, R., Gepstein, S. and Blumwald, E.** (2007). Delayed leaf senescence induces extreme drought tolerance in a flowering plant. *Proceedings of the National Academy of Sciences* **104**, 19631–19636.
- Rodrigues, J. A. and Zilberman, D.** (2015). Evolution and function of genomic imprinting in plants. *Genes & Development* **29**, 2517–2531.
- Rosado, A., Amaya, I., Valpuesta, V., Cuartero, J., Botella, M. A. and Borsani, O.** (2006). ABA- and ethylene-mediated responses in osmotically stressed tomato are regulated by the TSS2 and TOS1 loci. *Journal of Experimental Botany* **57**, 3327–3335.
- Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R. and Chivian, E.** (2001). Climate change and extreme weather events: Implications for food production, plant diseases and pests. *Global Change & Human Health* **2**, 90–104.
- Roulin, A., Auer, P. L., Libault, M., Schlueter, J., Farmer, A., May, G., Stacey, G., Doerge, R. W. and Jackson, S. A.** (2012). The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* **73**, 143–153.

- Rubinovich, L. and Weiss, D.** (2010). The Arabidopsis cysteine-rich protein GAS4 promotes GA responses and exhibits redox activity in bacteria and in planta. *The Plant Journal* **64**, 1018–1027.
- Ryu, H. and Cho, Y.-G.** (2015). Plant hormones in salt stress tolerance. *Journal of Plant Biology* **58**, 147–155.
- Saand, M. A., Xu, Y.-P., Munyampundu, J.-P., Li, W., Zhang, X.-R. and Cai, X.-Z.** (2015). Phylogeny and evolution of plant cyclic nucleotide-gated ion channel (CNGC) gene family and functional analyses of tomato CNGCs. *DNA Research* **22**, 471–483.
- Sairam, R. K., Rao, K. V. and Srivastava, G. C.** (2002). Differential response of wheat genotypes to long term salinity stress in relation to oxidative stress, antioxidant activity and osmolyte concentration. *Plant Science* **163**, 1037–1046.
- Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2004). Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. *Plant Physiology* **136**, 2734–2746.
- Sakuraba, Y., Jeong, J., Kang, M.-Y., Kim, J., Paek, N.-C. and Choi, G.** (2014a). Phytochrome-interacting transcription factors PIF4 and PIF5 induce leaf senescence in Arabidopsis. *Nature* **5**, 4636.
- Sakuraba, Y., Kim, D., Kim, Y.-S., Hörtensteiner, S. and Paek, N.-C.** (2014b). Arabidopsis STAYGREEN-LIKE (SGRL) promotes abiotic stress-induced leaf yellowing during vegetative growth. *FEBS Letters* **588**, 3830–3837.
- Salazar, C., Hernandez, C. and Pino, M. T.** (2015). Plant water stress: Associations between ethylene and abscisic acid response. *Chilean journal of agricultural research* **75**, 71–79.
- Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.-Y., Jamin, M., Cutler, S. R., Rodriguez, P. L. and Márquez, J. A.** (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* **462**, 665–668.
- Savage, R. S., Ghahramani, Z., Griffin, J. E., Kirk, P. and Wild, D. L.** (2013). Identifying cancer subtypes in glioblastoma by combining genomic, transcriptomic and epigenomic data . In *International Conference on Machine Learning (ICML) 2012: Workshop on Machine Learning in Genetics and Genomics.*, pp. 1–17.
- Schnable, J. C., Wang, X., Pires, C. J. and Freeling, M.** (2012). Escape from preferential retention following repeated whole genome duplications in plants. *Frontiers in Plant Science* **3**, 94.
- Schwechheimer, C. and Bevan, M.** (1998). The regulation of transcription factor activity in plants. *Trends in Plant Science* **3**, 378–383.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T. et al.** (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal* **31**, 279–292.
- Sengupta, A., Gupta, K. and Gupta, B.** (2015). Computational Biology and Chemistry. *Computational Biology and Chemistry* **54**, 18–32.

- Seo, P. J., Park, J.-M., Kang, S. K., Kim, S.-G. and Park, C.-M. (2010). An Arabidopsis senescence-associated protein SAG29 regulates cell viability under high salinity. *Planta* **233**, 189–200.
- Seo, P. J., Xiang, F., Qiao, M., Park, J. Y., Lee, Y. N., Kim, S. G., Lee, Y. H., Park, W. J. and Park, C. M. (2009). The MYB96 Transcription Factor Mediates Absciscic Acid Signaling during Drought Stress Response in Arabidopsis. *Plant Physiology* **151**, 275–289.
- Shabala, S. N. and Lew, R. R. (2002). Turgor Regulation in Osmotically Stressed Arabidopsis Epidermal Root Cells. Direct Support for the Role of Inorganic Ion Uptake as Revealed by Concurrent Flux and Cell Turgor Measurements. *Plant Physiology* **129**, 290–299.
- Shang, Y., Yan, L., Liu, Z.-Q., Cao, Z., Mei, C., Xin, Q., Wu, F.-Q., Wang, X.-F., Du, S.-Y., Jiang, T. et al. (2012). The Mg-Chelatase H Subunit of Arabidopsis Antagonizes a Group of WRKY Transcription Repressors to Relieve ABA-Responsive Genes of Inhibition. *The Plant Cell* **22**, 1909–1935.
- Sharma, P., Jha, A. B., Dubey, R. S. and Pessarakli, M. (2012). Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. *Journal of Botany* **2012**, 217037.
- Sharma, R., Mishra, M., Gupta, B., Parsania, C., Singla-Pareek, S. L. and Pareek, A. (2015). De Novo Assembly and Characterization of Stress Transcriptome in a Salinity-Tolerant Variety CS52 of Brassica juncea. *PLoS ONE* **10**, e0126783.
- Sharp, R. E. and LeNoble, M. E. (2002). ABA, ethylene and the control of shoot and root growth under water stress. *Journal of Experimental Botany* **53**, 33–37.
- Shavrukov, Y. (2013). Salt stress or salt shock: which genes are we studying? *Journal of Experimental Biology* **64**, 119–127.
- Shen, X., Wang, Z., Song, X., Xu, J., Jiang, C., Zhao, Y., Ma, C. and Zhang, H. (2014). Transcriptomic profiling revealed an important role of cell wall remodeling and ethylene signaling pathway during salt acclimation in Arabidopsis. *Plant Molecular Biology* **86**, 303–317.
- Shendure, J. (2008). The beginning of the end for microarrays? *Nature Methods* **5**, 585–587.
- Sheveleva, E., Chmara, W., Bohnert, H. J. and Jensen, R. G. (2002). Increased Salt and Drought Tolerance by D-Ononitol Production in Transgenic Nicotiana tabacum L. *Plant Physiology* **115**, 1211–1219.
- Shi, H., Ishitani, M., Kim, C. and Zhu, J. K. (2000). The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na⁺/H⁺ antiporter. *PNAS* **97**, 6896–6901.
- Shin, D., Koo, Y. D., Lee, J., Lee, H.-J., Baek, D., Lee, S., Cheon, C.-I., Kwak, S.-S., Lee, S. Y. and Yun, D.-J. (2004). Athb-12, a homeobox-leucine zipper domain protein from Arabidopsis thaliana, increases salt tolerance in yeast by regulating sodium exclusion. *Biochemical and Biophysical Research Communications* **323**, 534–540.
- Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Current Opinion in Plant Biology* **3**, 217–233.

- Shinozaki, K., Yamaguchi-Shinozaki, K. and Seki, M.** (2003). Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology* **6**, 410–417.
- Shrivastava, P. and Kumar, R.** (2015). Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi Journal of Biological Sciences* **22**, 123–131.
- Siddiqui, M., Al-Khaishany, M., Al-Qutami, M., Al-Whaibi, M., Grover, A., Ali, H., Al-Wahibi, M. and Bukhari, N.** (2015). Response of Different Genotypes of Faba Bean Plant to Drought Stress. *International Journal of Molecular Sciences* **16**, 10214–10227.
- Siddiqui, Z. S., Khan, M. A., Kim, B. G., Huang, J. S. and Kwon, T. R.** (2008). Physiological Responses of Brassica napus Genotypes to Combined Drought and Salt Stress. *Plant Stress* **2**, 78–83.
- Singh, N., Mishra, A. and Jha, B.** (2013). Over-expression of the Peroxisomal Ascorbate Peroxidase (SbpAPX) Gene Cloned from Halophyte Salicornia brachiata Confers Salt and Drought Stress Tolerance in Transgenic Tobacco. *Marine Biotechnology* **16**, 321–332.
- Singh, N., Mishra, A. and Jha, B.** (2014). Ectopic over-expression of peroxisomal ascorbate peroxidase (SbpAPX) gene confers salt stress tolerance in transgenic peanut (Arachis hypogaea). *Gene* **547**, 119–125.
- Sîrbu, A., Kerr, G., Crane, M. and Ruskin, H. J.** (2012). RNA-Seq vs Dual- and Single-Channel Microarray Data: Sensitivity Analysis for Differential Expression and Clustering. *PLoS ONE* **7**, e50986.
- Sivitz, A. B., Hermand, V., Curie, C. and Vert, G.** (2012). Arabidopsis bHLH100 and bHLH101 Control Iron Homeostasis via a FIT-Independent Pathway. *PLoS ONE* **7**, e44843.
- Skipper, E. S.** (2010). *Investigating the Genetic Control of Postharvest Shelf Life and Vitamin C content in Broccoli (Brassica oleracea var. italica)*. Ph.D. thesis, University of Warwick.
- Snyder, C. S., Bruulsema, T. W., Jensen, T. L. and Fixen, P. E.** (2009). Review of greenhouse gas emissions from crop production systems and fertilizer management effects. *Agriculture, Ecosystems & Environment* **133**, 247–266.
- So, H. A., Choi, S. J., Chung, E. and Lee, J. H.** (2015). Molecular characterization of stress-inducible PLATZ gene from soybean (Glycine max L.) . *Plant Omics Journal* **8**, 479–484.
- Soitamo, A. J., Piippo, M., Allahverdiyeva, Y., Battchikova, N. and Aro, E.-M.** (2008). Light has a specific role in modulating Arabidopsis gene expression at low temperature. *BMC Plant Biology* **8**, 13.
- Soltis, D. E., Visger, C. J. and Soltis, P. S.** (2014). The polyploidy revolution then...and now: Stebbins revisited. *American Journal of Botany* **101**, 1057–1078.
- Son, G. H., Wan, J., Kim, H. J., Nguyen, X. C., Chung, W. S., Hong, J. C. and Stacey, G.** (2011). Ethylene-Responsive Element-Binding Factor 5, ERF5, Is Involved in Chitin-Induced Innate Immunity Response. *MPMI* **25**, 48–60.

- Son, O., Hur, Y. S., Kim, Y. K., Lee, H. J., Kim, S. Y., Kim, M. R., Nam, K. H., Lee, M. S., Kim, B. Y., Park, J. et al. (2010). ATHB12, an ABA-Inducible Homeodomain-Leucine Zipper (HD-Zip) Protein of Arabidopsis, Negatively Regulates the Growth of the Inflorescence Stem by Decreasing the Expression of a Gibberellin 20-Oxidase Gene. *Plant and Cell Physiology* **51**, 1537–1547.
- Song, X., Li, Y. and Hou, X. (2013). Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinesecabbage (*Brassica rapa* ssp. *pekinensis*). *BMC Genomics* **14**, 573.
- Stegle, O., Denby, K. J., Cooke, E. J., Wild, D. L., Ghahramani, Z. and Borgwardt, K. M. (2010). A Robust Bayesian Two-Sample Test for Detecting Intervals of Differential Gene Expression in Microarray Time Series. *Journal of Computational Biology* **17**, 355–367.
- Steinhorst, L. and Kudla, J. (2013). Calcium and Reactive Oxygen Species Rule the Waves of Signaling. *Plant Physiology* **163**, 471–485.
- Stepien, P. and Johnson, G. N. (2008). Contrasting Responses of Photosynthesis to Salt Stress in the Glycophyte Arabidopsis and the Halophyte *Thellungiella*: Role of the Plastid Terminal Oxidase as an Alternative Electron Sink. *Plant Physiology* **149**, 1154–1165.
- Su, J., Wu, S., Xu, Z., Qiu, S., Luo, T., Yang, Y., Chen, Q., Xia, Y., Zou, S., Huang, B.-L. et al. (2013). Comparison of Salt Tolerance in Brassicas and Some Related Species. *American Journal of Plant Sciences* **04**, 1911–1917.
- Sun, Y., Kong, X., Li, C., Liu, Y. and Ding, Z. (2015). Potassium Retention under Salt Stress Is Associated with Natural Variation in Salinity Tolerance among Arabidopsis Accessions. *PLoS ONE* **10**, e0124032.
- Sunarpi, Horie, T., Motoda, J., Kubo, M., Yang, H., Yoda, K., Horie, R., Chan, W.-Y., Leung, H.-Y., Hattori, K. et al. (2005). Enhanced salt tolerance mediated by AtHKT1 transporter-induced Na⁺ unloading from xylem vessels to xylem parenchyma cells. *The Plant Journal* **44**, 928–938.
- Sunkar, R., Kapoor, A. and Zhu, J. K. (2006). Posttranscriptional Induction of Two Cu/Zn Superoxide Dismutase Genes in Arabidopsis Is Mediated by Downregulation of miR398 and Important for Oxidative Stress Tolerance. *The Plant Cell* **18**, 2051–2065.
- Suzuki, N., Bassil, E., Hamilton, J. S., Inupakutika, M. A., Zandalinas, S. I., Tripathy, D., Luo, Y., Dion, E., Fukui, G., Kumazaki, A. et al. (2016). ABA Is Required for Plant Acclimation to a Combination of Salt and Heat Stress. *PLoS ONE* **11**, e0147625.
- Suzuki, T., Matsuura, T., Kawakami, N. and Noda, K. (2000). Accumulation and leakage of abscisic acid during embryo development and seed dormancy in wheat. *Plant Growth Regulation* **30**, 253–260.
- Sze, H., Padmanaban, S., Cellier, F., Honys, D., Cheng, N. H., Bock, K. W., Conejero, G., Li, X., Twell, D., Ward, J. M. et al. (2004). Expression Patterns of a Novel AtCHX Gene Family Highlight Potential Roles in Osmotic Adjustment and K⁺ Homeostasis in Pollen Development. *Plant Physiology* **136**, 2532–2547.
- Taji, T., Seki, M., Yamaguchi-Shinozaki, K., Kamada, H., Giraudat, J. and Shinozaki, K. (1999). Mapping of 25 Drought-Inducible Genes, RD and ERD in Arabidopsis thaliana. *Plant Cell Physiology* **40**, 119–123.

- Tam, T. H. Y., Catarino, B. and Dolan, L.** (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proceedings of the National Academy of Sciences* **112**, E3959–E3968.
- Tang, H. and Lyons, E.** (2012). Unleashing the genome of *Brassica rapa*. *Frontiers in Plant Science* **3**, 172.
- Tao, J.-J., Chen, H.-W., Ma, B., Zhang, W.-K., Chen, S.-Y. and Zhang, J.-S.** (2015). The Role of Ethylene in Plants Under Salinity Stress. *Frontiers in Plant Science* **6**, 1059.
- Tapia, G., Morales-Quintana, L., Parra, C., Berbel, A. and Alcorta, M.** (2013). Study of nsLTPs in *Lotus japonicus* genome reveal a specific epidermal cell member (LjLTP10) regulated by drought stress in aerial organs with a putative role in cutin formation. *Plant Molecular Biology* **82**, 485–501.
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A.** (2011). Differential expression in RNA-seq: a matter of depth. *Genome Research* **21**, 2213–2223.
- Teige, M., Scheikl, E., Eulgem, T., Doczi, R., Ichimura, K., Shinozaki, K., Dangel, J. L. and Hirt, H.** (2004). The MKK2 Pathway Mediates Cold and Salt Stress Signaling in *Arabidopsis*. *Molecular Cell* **15**, 141–152.
- Tester, M.** (2015). ScienceDirect. *Procedia Environmental Sciences* **29**, 300–301.
- Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S. Y., Howe, G. A. and Browse, J.** (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature* **448**, 661–665.
- Thompson, A.** (2011). Dissecting QTL for water use efficiency in *Brassica oleracea*. In *UK Brassica Research Council*, pp. 1–12.
- Thompson, A. J.** (2009). Improving water use efficiency and water capture *Solanum* and *Brassica*, transgenic and QTL approaches. In *VeGIN Meeting*, pp. 1–10.
- Tiwari, M., Sharma, D., Singh, M., Tripathi, R. D. and Trivedi, P. K.** (2014). Expression of OsMATE1 and OsMATE2 alters development, stress responses and pathogen susceptibility in *Arabidopsis*. *Scientific Reports* **4**, 3964.
- Tiwari, S. B., Wang, X. J., Hagen, G. and Guilfoyle, T. J.** (2001). AUX/IAA Proteins Are Active Repressors, and Their Stability and Activity Are Modulated by Auxin. *The Plant Cell* **13**, 2809–2822.
- Toledo-Ortiz, G., Huq, E. and Quail, P. H.** (2003). The *Arabidopsis* Basic/Helix-Loop-Helix Transcription Factor Family. *The Plant Cell* **15**, 1749–1770.
- Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., Hine, E. E., Althoff, R., Arbogast, T. S., Tallon, L. J. et al.** (2006). Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell* **18**, 1348–1359.
- Tracy, F. E., Gilliam, M., Dodd, A. N., Webb, A. A. R. and Tester, M.** (2008). NaCl-induced changes in cytosolic free Ca²⁺ in *Arabidopsis thaliana* are heterogeneous and modified by external ionic composition. *Plant, Cell and Environment* **31**, 1063–1073.

- Traka, M. H., Saha, S., Huseby, S., Kopriva, S., Walley, P. G., Barker, G. C., Moore, J., Mero, G., van den Bosch, F., Constant, H. et al. (2013). Genetic regulation of glucoraphanin accumulation in Beneforté ®broccoli. *New Phytologist* **198**, 1085–1095.
- Tran, L. S. P., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2010). Role of cytokinin responsive two-component system in ABA and osmotic stress signalings . *Plant Signaling & Behavior* **5**, 148–150.
- Tran, L. S. P., Urao, T., Qin, F., Maruyaa, K., Kakimoto, T., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2007). Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. *Proceedings of the National Academy of Sciences* **104**, 20623–20628.
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L. (2012a). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* **31**, 46–53.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012b). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. and Pachter, L. (2010). nbt.1621. *Nature Biotechnology* **28**, 516–520.
- Trick, M., Cheung, F., Drou, N., Fraser, F., Lobenhofer, E. K., Hurban, P., Magusin, A., Town, C. D. and Bancroft, I. (2009). A newly-developed community microarray resource for transcriptome profiling in Brassica species enables the confirmation of Brassica-specific expressed sequences. *BMC Plant Biology* **9**, 50.
- Truernit, E., Siemering, K. R., Hodge, S., Grbic, V. and Haseloff, J. (2006). A Map of KNAT Gene Expression in the Arabidopsis Root. *Plant Molecular Biology* **60**, 1–20.
- U, N. (1935). Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. *Japan J. Bot* **7**, 389–452.
- Umar, S., Diva, I., Anjum, N., Iqbal, M., Ahmad, I. and Pereira, E. (2011). Potassium-induced alleviation of salinity stress in Brassica campestris L. *Central European Journal of Biology* **6**, 1054–1063.
- Vanderauwera, S., Vandenbroucke, K., Inze, A., van de Cotte, B., Muhlenbock, P., De Rycke, R., Naouar, N., Van Gaeve, T., Van Montagu, M. C. E. and Van Brusegem, F. (2012). AtWRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in Arabidopsis. *Proceedings of the National Academy of Sciences* **109**, 20113–20118.
- Vanneste, K., Maere, S. and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130353.
- Veitia, R. A. and Vaiman, D. (2011). Exploring the mechanistic bases of heterosis from the perspective of macromolecular complexes. *FASEB Journal* **25**, 476–482.
- Vermeulen, S. J., Campbell, B. M. and Ingram, J. S. I. (2012). Climate Change and Food Systems. *Annual Review of Environment and Resources* **37**, 195–222.

- Verslues, P. E.** (2016). ABA and cytokinins: challenge and opportunity for plant stress research. *Plant Molecular Biology* **91**, 629–640.
- Verslues, P. E. and Bray, E. A.** (2004). LWR1 and LWR2 Are Required for Osmoregulation and Osmotic Adjustment in Arabidopsis. *Plant Physiology* **136**, 2831–2842.
- Verslues, P. E. and Bray, E. A.** (2006). Role of abscisic acid (ABA) and Arabidopsis thaliana ABA-insensitive loci in low water potential-induced ABA and proline accumulation. *Journal of Experimental Botany* **57**, 201–212.
- Vincill, E. D., Bieck, A. M. and Spalding, E. P.** (2012). Ca²⁺ Conduction by an Amino Acid-Gated Ion Channel Related to Glutamate Receptors. *Plant Physiology* **159**, 40–46.
- Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M., Hernández-Oñate, M., Minoche, A. E., Erb, I., Câmara, F., Prieto-Barja, P., Corvelo, A., Sanseverino, W. et al.** (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biology* **17**, 32.
- Walia, H., Wilson, C., Condamine, P., LIU, X., Ismail, A. M. and Close, T. J.** (2007). Large-scale expression profiling and physiological characterization of jasmonic acid-mediated adaptation of barley to salinity stress. *Plant, Cell and Environment* **30**, 410–421.
- Walia, H., Wilson, C., Wahid, A., Condamine, P., Cui, X. and Close, T. J.** (2006a). Expression analysis of barley (*Hordeum vulgare* L.) during salinity stress. *Functional & Integrative Genomics* **6**, 143–156.
- Walia, H., Wilson, C., Zeng, L., Ismail, A. M., Condamine, P. and Close, T. J.** (2006b). Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Molecular Biology* **63**, 609–623.
- Walley, P. G., Carder, J., Skipper, E., Mathas, E., Lynn, J., Pink, D. and Buchanan-Wollaston, V.** (2011). A new broccoli broccoli immortal mapping population and framework genetic map: tools for breeders and complex trait analysis. *Theoretical and Applied Genetics* **124**, 467–484.
- Walley, P. G. and Moore, J. D.** (2015). Biotechnology and Genomics: Exploiting the Potential of CWR. In *Crop Wild Relatives and Climate Change* (eds. R. Redden, S. S. Yadav, N. Maxted, M. E. Dulloo, L. Guarino and P. Smith), pp. 212–223. John Wiley & Sons, Ltd.
- Walley, P. G., Teakle, G. R., Moore, J. D., Allender, C. J., Pink, D. A. C., Buchanan-Wollaston, V. and Barker, G. C.** (2012). Developing genetic resources for pre-breeding in *Brassica oleracea* L.: an overview of the UK perspective. *Journal of Plant Biotechnology* **39**, 62–68.
- Wang, H., Miyazaki, S., Kawai, K., Deyholos, M. K., Galbraith, D. and Bohnert, H. J.** (2003). Temporal progression of gene expression responses to salt shock in maize roots. *Plant Molecular Biology* **52**, 873–891.
- Wang, K. L. C., Li, H. and Ecker, J. R.** (2002). Ethylene Biosynthesis and Signaling Networks. *The Plant Cell* **14**, S131–S151.

- Wang, S., Bai, Y., Shen, C., Wu, Y., Zhang, S., Jiang, D., Guilfoyle, T. J., Chen, M. and Qi, Y. (2010). Auxin-related gene families in abiotic stress response in *Sorghum bicolor*. *Functional & Integrative Genomics* **10**, 533–546.
- Wang, S., Su, S. Z., Wu, Y., Li, S. P., Shan, X. H., Liu, H. K., Wang, S. and Yuan, Y. P. (2014). Overexpression of maize chloride channel gene ZmCLC-d in *Arabidopsis thaliana* improved its stress resistance. *Biologia Plantarum* **59**, 55–64.
- Wang, X., Torres, M. J., Pierce, G., Lemke, C., Nelson, L. K., Yuksel, B., Bowers, J. E., Marler, B. S., Xiao, Y., Lin, L. et al. (2011a). A physical map of *Brassica oleracea* shows complexity of chromosomal changes following recursive paleopolyploidizations. *BMC Genomics* **12**, 470.
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F. et al. (2011b). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature* **43**, 1035–1039.
- Wang, Y. and Hua, J. (2009). A moderate decrease in temperature induces COR15a expression through the CBF signaling cascade and enhances freezing tolerance. *The Plant Journal* **60**, 340–349.
- Wang, Y., Li, K. and Li, X. (2009a). Auxin redistribution modulates plastic development of root system architecture under salt stress in *Arabidopsis thaliana*. *Journal of Plant Physiology* **166**, 1637–1645.
- Wang, Y., Shen, W., Chan, Z. and Wu, Y. (2015). Endogenous Cytokinin Overproduction Modulates ROS Homeostasis and Decreases Salt Stress Resistance in *Arabidopsis Thaliana*. *Frontiers in Plant Science* **6**, 1004.
- Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S. P., Feltus, F. A. and Paterson, A. H. (2011c). Modes of Gene Duplication Contribute Differently to Genetic Novelty and Redundancy, but Show Parallels across Divergent Angiosperms. *PLoS ONE* **6**, e28150.
- Wang, Y., Yang, L., Zheng, Z., Grumet, R., Loescher, W., Zhu, J.-K., Yang, P., Hu, Y. and Chan, Z. (2013). Transcriptomic and Physiological Variations of Three *Arabidopsis* Ecotypes in Response to Salt Stress. *PLoS ONE* **8**, e69036.
- Wang, Z., Gerstein, M. and Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63.
- Ward, J. M. and Schroeder, J. I. (1994). Calcium-Activated K⁺ Channels and Calcium-Induced Calcium Release by Slow Vacuolar Ion Channels in Guard Cell Vacuoles Implicated in the Control of Stomatal Closure. *The Plant Cell* **6**, 669–683.
- Ward, J. M., Ward, J. M., Maser, P., Maser, P., Schroeder, J. I. and Schroeder, J. I. (2009). Plant Ion Channels: Gene Families, Physiology, and Functional Genomics Analyses. *Annual Review of Physiology* **71**, 59–82.
- Wasternack, C. and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. *Annals of Botany* **111**, 1021–1058.

- Wen, X., Zhang, C., Ji, Y., Zhao, Q., He, W., An, F., Jiang, L. and Guo, H. (2012). Activation of ethylene signaling is mediated by nuclear translocation of the cleaved EIN2 carboxyl terminus. *Cell Research* **22**, 1613–1616.
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Molecular Biology* **42**, 225–249.
- West, G., Inzé, D. and Beemster, G. T. S. (2004). Cell Cycle Modulation in the Response of the Primary Root of Arabidopsis to Salt Stress. *Plant Physiology* **135**, 1050–1058.
- Wheeler, T. and von Braun, J. (2013). Climate Change Impacts on Global Food Security. *Science* **341**, 508–513.
- White, P. J., Hammond, J. P., King, G. J., Bowen, H. C., Hayden, R. M., Meacham, M. C., Spracklen, W. P. and Broadley, M. R. (2010). Genetic analysis of potassium use efficiency in Brassica oleracea. *Annals of Botany* **105**, 1199–1210.
- Wilkins, O., Bräutigam, K. and Campbell, M. M. (2010). Time of day shapes Arabidopsis drought transcriptomes. *The Plant Journal* **63**, 715–727.
- Wilkinson, S. and Davies, W. J. (2002). ABA-based chemical signalling: the co-ordination of responses to stress in plants. *Plant, Cell and Environment* **25**, 195–210.
- Windram, O., Madhou, P., McHattie, S., Hill, C., Hickman, R., Cooke, E., Jenkins, D. J., Penfold, C. A., Baxter, L., Breeze, E. et al. (2012). Arabidopsis Defense against Botrytis cinerea: Chronology and Regulation Deciphered by High-Resolution Temporal Transcriptomic Analysis. *The Plant Cell* **24**, 3530–3557.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V. and Provart, N. J. (2007). An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE* **2**, e718.
- Won, J. H., Kim, W. and Hwang, B. K. (2003). Three pathogen-inducible genes encoding lipid transfer protein from pepper are differentially activated by pathogens, abiotic, and environmental stresses. *Plant, Cell and Environment* **26**, 915–928.
- Woodhouse, M. R., Cheng, F., Pires, C. J., Lisch, D., Freeling, M. and Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proceedings of the National Academy of Sciences* **111**, 6527–6527.
- Wortman, J. R., Haas, B. J., Hannick, L. I., Smith, R. K., Maiti, R., Ronning, C. M., Chan, A. P., Yu, C., Ayele, M., Whitelaw, C. A. et al. (2003). Annotation of the Arabidopsis Genome. *Plant Physiology* **132**, 461–468.
- Wu, H., Kerr, K., Cui, X. and Churchill, G. A. (2002). *MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments*.
- Wu, L.-t., Zhong, G.-m., Wang, J.-m., Li, X.-f. and Yang, Y. (2011). Arabidopsis WRKY28 transcription factor is required for resistance to necrotrophic pathogen, Botrytis cinerea. *African Journal of Microbiology Research* **5**, 5481–5488.

- Wu, S. J., Ding, L. and Zhu, J. K. (1996). SOS1, a Genetic Locus Essential for Salt Tolerance and Potassium Acquisition. *The Plant Cell* **8**, 617–627.
- Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., Zhao, Q., Zhang, L., Li, Y., Yang, C. et al. (2009). Dual function of *Arabidopsis* ATAF1 in abiotic and biotic stress responses. *Cell Research* **19**, 1279–1290.
- Xu, P., Liu, Z., Fan, X., Gao, J., Zhang, X., Zhang, X. and Shen, X. (2013a). *De novo* transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. *Gene* **525**, 26–34.
- Xu, Y.-X., Mao, J., Chen, W., Qian, T.-T., Liu, S.-C., Hao, W.-J., Li, C.-F. and Chen, L. (2016). Plant Physiology and Biochemistry. *Plant Physiology et Biochemistry* **98**, 46–56.
- Xu, Z. Y., Kim, S. Y., Hyeon, D. Y., Kim, D. H., Dong, T., Park, Y., Jin, J. B., Joo, S. H., Kim, S. K., Hong, J. C. et al. (2013b). The Arabidopsis NAC Transcription Factor ANAC096 Cooperates with bZIP-Type Transcription Factors in Dehydration and Osmotic Stress Responses. *The Plant Cell* **25**, 4708–4724.
- Xue, S., Yao, X., Luo, W., Jha, D., Tester, M., Horie, T. and Schroeder, J. I. (2011). AtHKT1;1 Mediates Nernstian Sodium Channel Transport Properties in Arabidopsis Root Stelar Cells. *PLoS ONE* **6**, e24725.
- Yamaguchi-Shinozaki, K. and Shinozaki, K. (1993). The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in Arabidopsis thaliana. *Molecular Genetics and Genomics* **238**, 17–25.
- Yang, O., Popova, O. V., Süthoff, U., Lüking, I., Dietz, K.-J. and Golldack, D. (2009). The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene* **436**, 45–55.
- Yang, Z., Liu, J., Tischer, S. V., Christmann, A., Windisch, W., Schnyder, H. and Grill, E. (2016). Leveraging abscisic acid receptors for efficient water use in Arabidopsis. *Proceedings of the National Academy of Sciences* **113**, 6791–6796.
- Yang, Z., Tian, L., Latoszek-Green, M., Brown, D. and Wu, K. (2005). Arabidopsis ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses. *Plant Molecular Biology* **58**, 585–596.
- Yoo, J. H., Park, C. Y., Kim, J. C., Do Heo, W., Cheong, M. S., Park, H. C., Kim, M. C., Moon, B. C., Choi, M. S., Kang, Y. H. et al. (2005). Direct Interaction of a Divergent CaM Isoform and the Transcription Factor, MYB2, Enhances Salt Tolerance in Arabidopsis. *Journal of Biological Chemistry* **280**, 3697–3706.
- Yoon, J. Y., Hamayun, M., Lee, S.-K. and Lee, I.-J. (2009). Methyl jasmonate alleviated salinity stress in soybean. *Journal of Crop Science and Biotechnology* **12**, 63–68.
- Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., TODAKA, D., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2014a). Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. *Plant, Cell and Environment* **38**, 35–49.

- Yoshida, T., Fujita, Y., Sayama, H., Kidokoro, S., Maruyama, K., Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2010). AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *The Plant Journal* **61**, 672–685.
- Yoshida, T., Mogami, J. and Yamaguchi-Shinozaki, K.** (2014b). ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Current Opinion in Plant Biology* **21**, 133–139.
- Yubero-Serrano, E.-M., Moyano, E., Medina-Escobar, N., Munoz-Blanco, J. and Caballero, J.-L.** (2003). Identification of a strawberry gene encoding a non-specific lipid transfer protein that responds to ABA, wounding and cold stress. *Journal of Experimental Botany* **54**, 1865–1877.
- Zhai, H., Wang, F., Si, Z., Huo, J., Xing, L., An, Y., He, S. and Liu, Q.** (2015). A myo-inositol-1-phosphate synthase gene, IbMIPS1, enhances salt and drought tolerance and stem nematode resistance in transgenic sweet potato. *Plant Biotechnology Journal* **14**, 592–602.
- Zhang, J., Jia, W., Yang, J. and Ismail, A. M.** (2006). Role of ABA in integrating plant responses to drought and salt stresses. *Field Crops Research* **97**, 111–119.
- Zhang, M., Yuan, B. and Leng, P.** (2009). The role of ABA in triggering ethylene biosynthesis and ripening of tomato fruit. *Journal of Experimental Botany* **60**, 1579–1588.
- Zhang, X., Liu, X., Wu, L., Yu, G., Wang, X. and Ma, H.** (2015). The SsDREB Transcription Factor from the Succulent Halophyte Suaeda salsa Enhances Abiotic Stress Tolerance in Transgenic Tobacco. *International Journal of Genomics* **2015**, 875497.
- Zhang, Y., Liu, Z., Khan, M. A., Lin, Q., Han, Y., Mu, P., Liu, Y., Zhang, H., Li, L., Meng, X. et al.** (2016). Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.). *Nature* **6**, 21476.
- Zhao, J. and Dixon, R. A.** (2009). MATE Transporters Facilitate Vacuolar Uptake of Epicatechin 3'-O-Glucoside for Proanthocyanidin Biosynthesis in Medicago truncatula and Arabidopsis. *The Plant Cell* **21**, 2323–2340.
- Zhou, S., Zhang, Z., Tang, Q., Lan, H., Li, Y. and Luo, P.** (2010). Enhanced V-ATPase activity contributes to the improved salt tolerance of transgenic tobacco plants overexpressing vacuolar Na⁺/H⁺ antiporter AtNHX1. *Biotechnology Letters* **33**, 375–380.
- Zhu, J., Fu, X., Koo, Y. D., Zhu, J. K., Jenney, F. E., Adams, M. W. W., Zhu, Y., Shi, H., Yun, D. J., Hasegawa, P. M. et al.** (2007). An Enhancer Mutant of Arabidopsis salt overly sensitive 3 Mediates both Ion Homeostasis and the Oxidative Stress Response. *Molecular and Cellular Biology* **27**, 5214–5224.
- Zhu, J. K., Liu, J. and Xiong, L.** (1998). Genetic Analysis of Salt Tolerance in Arabidopsis: Evidence for a Critical Role of Potassium Nutrition. *The Plant Cell* **10**, 1181–1191.